BMJ Open

ULTRASONIC STRAIN ELASTOGRAPHY FOR DETECTING ABNORMALITIES WITHIN THE SUPRASPINATUS TENDON: AN INTRA- AND INTERRATER RELIABILITY STUDY

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-027725
Article Type:	Research
Date Submitted by the Author:	08-Nov-2018
Complete List of Authors:	Brage, K; University of Southern Denmark, Sport Science and Clinical Biomechanics Hjarbaek, John; Odense Universitetshospital, Department of Radiology, Musculoskeletal section Kjaer, Per ; University of Southern Denmark, Institute of Sports Science and Clinical Biomechanics Ingwersen, Kim G.; Sports Science and Clinical Biomechanics Juul-Kristensen, Birgit; Research Unit for Musculoskeletal Function and Physiotherapy, Institute of Sports Science and Clinical Biomechanics
Keywords:	Strain elastography, supraspinatus tendon, rotator cuff, tendon quality, reliability, tendinopathy

ULTRASONIC STRAIN ELASTOGRAPHY FOR DETECTING ABNORMALITIES WITHIN THE SUPRASPINATUS TENDON: AN INTRA- AND INTERRATER RELIABILITY STUDY

Brage K[1,2], Hjarbaek J[3], Kjaer P[1,2], Ingwersen KG[4], Juul-Kristensen B[1]

[1] Department of Sports Science and Clinical Biomechanics, University of Southern Denmark, Campusvej 55, 5230 Odense M, DK
[2] UCL University College, Niels Bohrs Allé 1, 5230, Odense M, DK
[3] Department of Radiology, Odense University Hospital, J. B. Winsløws Vej 4, 5000 Odense C, DK
[4]Research Unit in Physio- and Occupational Therapy, Hospital Lillebaelt, Kabbeltoft 25, 7100 Vejle, DK

Author information
Corresponding Author
Karen Brage, PhD Fellow, Telephone: +45 51787381, fax: 6318 3226, e-mail: kabr2@ucl.dk
Co-Authors
John Hjarbaek, MD, Telephone: +45 66113333, e-mail: john.hjarbaek@rsyd.dk
Per Kjær, Professor, Telephone: +45 65504553, e-mail: pkjaer@health.sdu.dk
Kim Gordon Ingwersen, PhD., Telephone: +45 65509044, e-mail: kingwersen@health.sdu.dk
Birgit Juul-Kristensen, PhD.: Telephone: +45 65503412, e-mail: bjuul-kristensen@health.sdu.dk

Wordcount
4.224

Trial registration
The study protocol was approved by the Ethical Committee for Region South Denmark (S-20160115) and reported to the Danish Data Protection Agency (2014-41-3266).

ABSTRACT

Objectives

Reliability of strain-elastography (SEL) used for supraspinatus pathologies is unclear. Thus the aim was to investigate the reliability of SEL within the supraspinatus tendon.

Design

An intra- and inter-rater reliability study.

Setting

A single center study conducted at the University.

Participants

Twenty participants with shoulder pain and MRI-verified supraspinatus tendinosis and 20 asymptomatic participants (no MRI).

Primary and secondary outcome measures

Raw values (RAW), and ratios (deltoid muscle, DELT; gel pad, GEL, as reference tissues) were calculated and mean values of measurements from three regions of the supraspinatus tendon were reported. Color scale ratings and number of yellow/red lesions from the three areas were furthermore included.

Results

Intra-rater reliability showed Intra-Class Correlation Coefficients (ICC) for RAW, DELT and GEL at 0.97 (Minimal Detectable Change (MDC): 0.28 (6.36% of the mean), 0.89 (MDC: 2.91 (20.37%), and 0.73 (MDC: 1.61 (58.82%)), respectively. ICC for Inter-rater reliability were 0.89 (MDC: 0.47 (10.53%), 0.78 (MDC: 3.69 (25.51%), and 0.70 (MDC: 1.75 (62.63%)), respectively.

For color scale ratings, intra-rater reliability (Linear Weighted kappa, LWk) ranged from 0.76-0.79, with the inter-rater reliability from 0.71-0.81. For number of lesions intra-rater reliability ranged from 0.40-0.82, and inter-rater reliability from 0.24-0.67.

Conclusions

Intra- and inter-rater reliability was excellent for raw values and for ratios with deltoid as reference tissue, and good for ratios with gel pad as reference. Reliability of color scale ratings was substantial to almost perfect, and for number of lesions fair to almost perfect.

Although high reliability was found, validity and responsiveness of these elastographic methods must be established.

Strengths and limitations of this study

- A standardized procedure for capturing strain elastographic images, to measure pathology in the supraspinatus tendon, is presented.
- A specific procedure for grading strain elastographic images is presented.
- Validity of this elastographic method must be established.

INTRODUCTION

Shoulder pain is a common symptom, with a lifetime prevalence in the general population of 6.7-66.7%,[1] and subacromial pain syndrome is the second most common cause of pain in the shoulder,[2]. Shoulder pain have consequences such as physical limitations, mental problems,[3] and absence from work,[4].

Shoulder disorders are evaluated by history taking and physical exam, potentially supplemented with X-ray, conventional ultrasound (US) and/or Magnetic Resonance Imaging (MRI). However, supraspinatus tendon abnormalities are also found in asymptomatic people when using general modalities such as MRI and US,[5]. Furthermore, it seems difficult to distinguish pathological changes from healthy tissues by using conventional ultrasound because pathological regions often exhibit the same echogenicity as non-pathological regions,[6].

Strain elastography (SEL) is a relatively new and not yet well-established method, which may assist in early diagnosis, prediction and monitoring of progress in tendon healing,[7].

SEL defines the physical properties of soft tissues through characterization of the differences in stiffness between 'the regions of interest' (ROI) and the surrounding tissues,[8,9] Conventional US and MRI are developed for visualization of macroscopic changes and not specifically the mechanical tendon properties, why SEL may add further knowledge to conventional shoulder imaging. Since tendon quality is a prognostic factor for rotator cuff repair, information about tendon stiffness could be beneficial for the surgeon,[10].

Tissue deformation using SEL is obtained by uniformly mechanically induced compressions (strain) of the structures under the US transducer, during the US-scan. Through manual compression, the

soft tissue deforms differently depending on the inherent stiffness. The degree of deformation can be interpreted as an estimate of the tissue stiffness,[11].

SEL has recently been used in the musculoskeletal area, where the achilles tendon has been the primary focus,[12]. A few studies have found significant associations between pathology identified by SEL and MRI in patients with supraspinatus tendinopathy,[13, 14], besides significant correlations between results from SEL and clinical tests and questionnaires in patients with small supraspinatus tendon tears,[15]. One study also found SEL to be able to detect increased stiffness in the supraspinatus tendon- and muscle elasticity with increased muscle contraction in healthy participants [16]. Although only few studies have been performed, and with different aims, comparator instruments, procedures, reference tissues and data types, the validity of SEL in the supraspinatus tendon seems promising,[13-16].

However, firstly the reliability of SEL must be proven to be acceptable. In this respect SEL constitutes several challenges since SEL is a technique with relatively high operator dependency in terms of the manually applied pressure and afterwards the identification and selection of the pathological region of interest (ROI).

One of only two studies on reliability found the inter-rater reliability of SEL in the supraspinatus tendon to be almost perfect ($\kappa = 0.83$) with respect to number of focal lesions in 118 patients with MRI-verified supraspinatus tendinopathy. However, this study did not include a healthy control group and used only color quantification,[14]. The other study found a high intra-rater reliability ($ICC_{1,3} = 0.92\text{-}0.99$) with respect to ROI, when using an acoustic coupler as reference tissue in a small sample of 23 healthy participants. Limitations of this study were; not including a group with pathology, not defining ROI and only using one type of reference tissue (acoustic coupler),[16].

To our knowledge no study, investigating the reliability of SEL in the supraspinatus tendon, has included both patients with a pathological (non-ruptured) supraspinatus tendon and healthy participants with non-painful shoulders. Furthermore choice of reference tissue and quantification methods has major impact on results, but no reliability studies have compared these different approaches.

The aim of this study was to test the intra- and inter-rater reliability of SEL within the supraspinatus tendon in patients and healthy participants, using different reference tissues (deltoid muscle, gel pad), and different quantification methods (raw data, strain ratios, color scale rating and counting number of yellow/red lesions).

MATERIALS AND METHODS

This study is an intra- and inter-rater reliability study of SEL, used on the supraspinatus tendon, reported according to guidelines for reporting reliability and agreement studies (GRRAS),[17].

A 3-phased intra- and inter-rater reliability protocol for diagnostic reliability studies was used,[18]. The protocol included a training phase, (phase one), an overall agreement phase, (phase two), and an actual study phase, (phase three), for securing low clinician dependency and subjectivity, sufficient experience and standardization, and minimization of systematic bias,[18]. In order to continue to the study phase the criteria of at least 80% inter-rater agreement in phase two was used. Such reliability protocols have previously been used in reliability studies, using ultrasound methods,[19-20].

Study procedures

Study participants were recruited from August 2016 to December 2017. Symptomatic participants, with an MRI-verified supraspinatus tendinosis (*patients)*, were recruited from the Radiology Department at Odense University Hospital within 14 days after MRI examination. Participants with no shoulder symptoms (*healthy participants),* were recruited primarily from the University of Southern Denmark and University College of Lillebaelt, both in Odense. Except for MRI, all procedures were performed at University of Southern Denmark. After inclusion, participants underwent clinical tests (performed by KGI) and filled out questionnaires regarding functional limitations, pain and quality of life. Testing procedures took place once and lasted approximately one hour.

All SEL images were captured by the same clinician (KB), while two raters performed the SEL measurements on the captured images. Rater 1 (KB), radiographer, was thoroughly trained in using SEL on the shoulder by rater 2 (JH), radiologist, who have more than 20 years' of experience in clinical musculoskeletal US. In the study phase (phase 3), raters were blinded to each other's results (data was stored separately), participant's health status (Rater 1 entered the room after clinical tests and questionnaires were performed, and Rater 2 had no contact with the participants) and MRI results (ID-numbers were changed after MRI examinations).

SEL images were stored for at least 14 days after image capturing until the first image assessment, to ensure elimination of any memory of pain response during SEL by Rater 1. Further, all SEL images were stored for additional 14 days before rater 1's second assessment, to ensure elimination of recalls of SEL results from the first assessment.

The study protocol was approved by the Ethical Committee for Region South Denmark (S-20160115) and reported to the Danish Data Protection Agency (2014-41-3266). All participants had oral and written information about the study and signed an informed consent form before participation in this study.

Patient and Public Involvement

Patients and public were not involved in this study.

Participants

In- and exclusion criteria

Inclusion criteria for patients and healthy participants were age 18-65.

Patients had to have at least one positive out of the following five clinical tests: 'full can test', 'Jobe's test', 'resisted external rotation test', Hawkins-Kennedy test 'and 'Neer's test'. Furthermore, at least one shoulder had to be diagnosed with tendinosis based on MRI (≥ grade 1),[21]. If both shoulders were MRI-scanned the most severely affected side was included. Inclusion criteria for the healthy participants were no previously experienced shoulder problems, and negative signs from all five clinical tests described above. By convenience 50% of participants had their dominant arm scanned with strain elastography, while the remaining 50% had their non-dominant arm scanned.

Exclusion criteria for both groups were: Tears >1/3 of the vertical height of the supraspinatus tendon since the stress may be increased on the intact tendon part and calcifications >2 mm (length) due to acoustic shadowing. Further exclusion criteria were: previous comorbidities (potentially harmful to the tendon) such as; past/present shoulder fractures, operation and luxation, known neuromuscular disease, rheumatoid arthritis, cancer, fibromyalgia, spondyloarthropathy and psychiatric disorders. Pregnancy and inability to read and understand Danish were exclusion criteria as well.

SEL image capturing and measurement

Apparatus

All measurements were performed with the Logiq S7 using a 15 MHz linear probe (GE Healthcare, Milwaukee, USA). Manufacturer recommendations for musculoskeletal SEL of the shoulder were used, including a transducer movement rate of approximately 120 cycles/min, axial smoothing of 2, lateral smoothing of 3, frequenzy of 10 and a soft/hard compression of 5.

Patient placement

The SEL was obtained with the patient sitting in the erect position with the arm internally rotated, elbow flexed to 90° and with the dorsal side of the hand placed over the sacrum, as previously used [15]. The probe was placed on the anterior aspect of the acromion in a coronal plane and the images were obtained just laterally to the anterior-lateral corner of the acromion in the longitudinal plane of the supraspinatus tendon (Figure 1).

Image capturing

An image window depth of at least three times the tendon size and an image width covering about three-quarters of the screen were used as recommended for longitudinal SEL,[22]. The tissue was compressed approximately 2-5 mm [9], and a software incorporated quality control (expressed as 1-5 green bars being displayed, with 5 bars being the most acceptable) was used to evaluate the recommended compression size.

For each assessment method (with or without a gel pad covering the transducer (Sonar Aid, 10 mm; Geistlich Pharma,Wolhusen, Switzerland)) 3 sessions of 20 sec. were obtained.

Image measurements

Tendon characteristics were evaluated quantitatively and qualitatively.

Quantitatively, ROIs on the SEL images were drawn over the target area (supraspinatus tendon) and the exact raw strain value (RAW) (0-6; 6 being the hardest tissue) was calculated.

Further, an adjacent reference region (normal tissue, experiencing the same stress as the target region) was drawn. From these two variables the strain ratio (0-60; 60 being the hardest tissue) was calculated,[23]. Two different reference tissues were used; a 1 mm circle region in a soft part of the

deltoid/bursal area for the DELT measurement (Figure 2), and a 5 mm circle region in a gel pad covering the transducer, for the GEL measurement (Figure 3),[13, 24]. The gel pad was used as a more homogeneous reference tissue

For the raw values alone and the strain ratios (strain raw value of supraspinatus tendon (A) to strain raw value of reference tissue (B) = A/B), a mean of the three measured areas of the supraspinatus tendon was calculated including data from 5-15 sec. of the 20 sec. cycle as recommended by the manufacturer.

Quantitative measurements were based on examination of 3 entire cine-loops (10 sec) rather than on single static images,[25], in order to minimize intra-observer variation and avoid transient temporal fluctuations. Only sequences with the highest image quality (with green bars on the quality assessment) were used as recommended by the manufacturer.

Due to difficulties in defining the most lateral part of the supraspinatus insertion on the humeral head, a 6.5 mm chord was drawn (in the lateral direction) from the medial part of the insertion on the humeral head to the lateral part of the tuberculum major. 6.5 mm has previously been estimated to be the average length of the supraspinatus insertion,[26]. This fix point (end of the line at the lateral part of the insertion) was used to draw a 23 mm (7.7 x 3) horizontal line in the medial direction (which has been estimated to be the average length of the tendon) [27], ensuring agreement of measurement area (Figure 4).

Caudal borders for image measurements were bony surfaces (of the humeral head), while cranial borders were the transition zone between the (superior surface of the supraspinatus) tendon and the inferior surface of the deltoid muscle and bursa. The bursa area was used as reference tissues (red

color in the SEL image). If no red border was seen, an estimated border followed the superior surface of the tendon.

Qualitatively the images were rated using a color scale, from 1-4, according to the following: type 1, < 10% other color than blue (indicating a high tissue stiffness); type 2, 11-25% other color than blue; type 3, 26-50% other color than blue; or type 4, > 50% other color than blue (indicating a low tissue stiffness). Furthermore the number of yellow/red lesions was graded as follows; 0 = no lesions; 1 = one lesion; 2 = two lesions; and 3 = more than two lesions [14]. The qualitative classifications were all performed on the first high quality image recorded closest to 10 sec. into the first cycle.

Development of SEL method

Based on previous studies,[14, 16, 25, 28-31] a protocol with standardized procedures was developed for obtaining of SEL characteristics in the Supraspinatus tendon and tested in phase one on 10 participants.

Based on the results from these participants, adjustments concerning type of reference tissues and color scale criteria were performed. One adjustment was replacement of the subcutaneous fat with the deltoid muscle as reference tissue, since some participants had too thin subcutaneous fat layer to measure. Another adjustment was to base the color scale on percentage of colors (replacement of estimation of the most pronounced color), which made it possible to rate a blue (hard) tendon as softer if it appeared with yellow/red lesions.

In phase two the adjusted protocol was applied on 20 new participants. Overall agreement in phase two corresponded to >80% with blinded raters. Hereafter, phase three (the actual study phase, n=40) was initiated based on the final protocol, as described above.

Questionnaires

Participants completed questionnaires, including DASH (Disabilities of the Arm, Shoulder and Hand), for investigating disability of the upper extremities (0-100; 100 being most disabled),[32], VAS (Visual Analogue Scale), for assessing pain level (0-100; 100 being the most painful),[33], EQ-5D-3L, for measuring health-related quality of life and EQ-VAS also for health-related quality of life (0-100; 100 being best imaginable health status) [34]. Demographic data included information on age, gender, and BMI.

Statistics

SEL data was found to be normally distributed on a histogram with a normality distribution curve. For continuous data (mean of the three measured areas of the supraspinatus tendon) the intra-class correlation coefficient (ICC model 2.3, absolute-agreement, 2-way random, single measures) was calculated to determine intra- and inter-rater reliability,[35]. The ICC, with 95% confidence intervals (95% C.I.), was calculated for RAW, and for strain ratios (DELT and GEL). ICC were interpreted as <0.40 = poor, 0.40 to 0.59 = fair, 0.60-0.74 = good and ≥ 0.75 = excellent reliability [36].

Paired students t-tests were completed for statistical comparison of differences between raters (rater 1 (first time) vs. rater 1 (second time), and rater 1 vs. rater 2) using a significance level of 0.05. Bland and Altman plots with 95% limits of agreement (LOA) were calculated to evaluate systematic differences [37], with the 95% LOA calculated as the mean difference ± 1.96 x standard deviation of the difference (SD),[38].

Standard Error of Measurements (SEM) was calculated using the formula: SEM = SDmean difference/√2,[39] and Minimal Detectable Change 95% ($MDC_{individual}$) was calculated using the formula: MDC =SEM x √2 x 1.96 [40] and the relative MDC was calculated as a percentage of the average SEL value of Rater 1 and Rater 2.

For ordinal data linear weighted Cohen's κ (LWk) was used to calculate reliability with 95% CIs for color ratings and number of lesions. LWk was interpreted as <0=poor, 0.01 to 0.20=slight agreement, 0.21-0.40=fair agreement, 0.41 to 0.60=moderate agreement, 0.61 to 0.80=substantial agreement and >0.81=almost perfect agreement,[41].

Statistical analyses were performed in SPSS, version 25.0 (SPSS Inc., Chicago, IL).

RESULTS

Demographics varied between patients and healthy participants, on most parameters. Patients were, as expected due to the sampling method, older, had a higher BMI, more pain and disability, and reduced quality of life compared with healthy participants (Table 1).

Table 1
Demographics (mean; SD (frequencys for EQ-5D)) of patients and healthy participants from the study phase (n = 40).

Variable	**Patients (n = 20)**	**Healthy participants (n = 20)**
Age (years)	47.85 (7.63)	25.70 (6.10)
Gender (n (females (%))	14 (70)	11 (55)
BMI	30.49 (6.57)	24.94 (2.45)
VAS rest (0-100)	24.60 (20.96)	0 (0.00)
VAS activity (0-100)	53.35 (16.62)	0 (0.00)
VAS sleep (0-100)	49.60 (18.26)	0 (0.00)
VAS maximum (0-100)	78.05 (11.62)	0 (0.00)
DASH (0-100)	34.75 (17.48)	1.71 (4.21)
EQ-VAS (0-100)	53.24 (38.31)	72.90 (36.40)
EQ-5D-3L	**Frequencys**	**Frequencys**
Mobility problems	0	0
Self-care problems	10	1
Usual activities problems	1	0
Pain/discomfort problems	20	1
Anxiety/depression problems	5	1

Abbreviations: SD, Standard Deviation; BMI, Body Mass Index; VAS, Visual Analogue Scale; DASH, Disability of the Arm, Shoulder and Hand Questionnaire; EQ-5D-3L, Quality of Life by dimension; EQ-VAS, Quality of Life

The paired t-test showed statistical differences in inter-rater measurements for RAW (Table 2) and in intra-rater measurements for GEL.

Table 2
Reliability of strain elastography in the supraspinatus tendon using respectively the deltoid muscle (reference) and a gelpad (reference) and raw data from the study phase (n = 40).

Continuous scale	Mean (SD)	Mean (SD)	Mean Difference (SD)	P	LOA	SEM	MDC (%)	ICC (95% C.I.)
Mean across tendon								
Intra-rater	**Rater 1**	**Rater 1**						
RAW	4.40 (0.55)	4.36 (0.55)	0.04 (0.14)	0.09	-0.24; 0.33	0.10	0.28 (6.36)	0.97 (0.93-0.98)
DELT (ratio)	14.33 (3.07)	14.23 (3.16)	0.10 (1.48)	0.66	-2.81; 3.01	1.05	2.91 (20.37)	0.89 (0.80-0.94)
GEL (ratio)	2.87 (1.28)	2.59 (1.01)	0.28 (0.82)	0.04*	-1.33; 1.89	0.58	1.61 (58.82)	0.73 (0.54-0.85)
Inter-rater	**Rater 1**	**Rater 2**						
RAW	4.40 (0.55)	4.53 (0.61)	-0.13 (0.24)	0.00*	-0.60; 0.34	0.17	0.47 (10.53)	0.89 (0.75-0.95)
DELT (ratio)	14.33 (3.07)	14.56 (2.60)	-0.22 (1.88)	0.46	-3.91; 3.46	1.33	3.69 (25.51)	0.78 (0.63-0.88)
GEL (ratio)	2.87 (1.28)	2.73 (1.00)	0.14 (0.89)	0.32	-1.61; 1.90	0.63	1.75 (62.63)	0.70 (0.50-0.83)

Abbreviations: SD, Standard Deviation; LOA, Limits of Agreement; SEM, Standard Error of Mean; MDC, Minimal Detectable Change; ICC (95% C.I.), Intra-class Correlation Coefficient with 95% confidence intervals; DELT, Reference area in the Deltoid Muscle; GEL, Reference area in the Gel Pad; Raw, Raw Elastography Data; *, significant difference ($p \leq 0.05$) between first and second measurements for Rater 1.

None of the Bland and Altman plots showed funnel effects (increasing difference with increasing mean size) (Figure 5).

For intra-rater and inter-rater reliability, there was 'excellent agreement' using RAW and DELT (ICC: intra-rater: 0.97, inter-rater: 0.89), (ICC: intra-rater: 0.89, inter-rater: 0.78) respectively and the reliability was 'good' when using GEL (ICC: intra-rater: 0.73, inter-rater: 0.70). (Table 2).

The same pattern, for all measurements, was seen for each of the three measured areas of the supraspinatus tendon (not shown in tables).

For the intra-rater reliability the relative MDC was smallest for RAW (6.36%), larger for DELT (20.37%) and largest for GEL (58.82%). For inter-rater reliability the same pattern was seen for MDC, with a minimum of 10.53%, 25.51% and 62.63%, for RAW, DELT and GEL, respectively.

Reliability of using the color scale (performed without gel pad) kappa (LWk) was very similar, with intra-rater reliability ranging from LWk: 0.76-0.79, representing 'substantial agreement', and inter-rater agreement ranging from LWk: 0.71-0.81 representing 'substantial to almost perfect agreement'.

For the number of yellow/red lesions, LWk was generally highest for intra-rater ranging from 0.40-0.82 representing 'fair to almost perfect agreement', while for inter-rater reliability LWk ranged from 0.24-0.67 representing 'fair to substantial agreement' (Table 3).

Table 3

Reliability of strain elastography in the supraspinatus tendon using data based on colors, respectively a colorscale and counting no. (number) of lesions from the study phase (n=40)

Ordinal scale	**Intra-rater**		**Inter-rater**	
	Total Agreement (%)	LWκ (95% C.I.)	Total Agreement (%)	LWκ (95% C.I.)
Colorscale				
Medial tendon 1/3	80	0.76 (0.64-0.89)	80	0.77 (0.63-0.91)
Middle tendon 1/3	82.5	0.77 (0.65-0.88)	76	0.71 (0.58-0.83)
Lateral tendon 1/3	85	0.79 (0.65-0.94)	87.5	0.81 (0.65-0.97)
Lesions (no.)				
Medial tendon 1/3	87.5	0.82 (0.69-0.95)	82.5	0.67 (0.53-0.80)
Middle tendon 1/3	90	0.75 (0.51-1.00)	87.5	0.63 (0.30-0.96)
Lateral tendon 1/3	87.5	0.40 (-0.16-0.97)	90	0.24 (-0.14-0.62)

Abbreviations: LWκ (95% C.I.), Linear Weighted κ with 95% confidence intervals

For the medial and middle tendon part LWk was ‘substantial to almost perfect’ for intra-rater (LWk 0.75-0.82) and ‘substantial’ for inter-rater (LWk 0.63-0.67) reliability, while for the lateral tendon part LWk was low, corresponding to only 0.40 and 0.24 for intra-rater and inter-rater reliability, respectively, representing ‘fair agreement’.

DISCUSSION

The reliability of the results from SEL was 'excellent' when using the raw data and the deltoid muscle as reference tissue. When using a gel pad as reference the reliability of the results was 'good'. Furthermore, the relative MDC (% of mean) was smallest for RAW and largest for GEL.

When using the color scale grading LWk represented 'substantial agreement' to 'substantial to almost perfect agreement'. For the number of yellow/red lesions, LWK was highest for intra-rater reliabiliy, in the medial and middle tendon part LWk was 'substantial to almost perfect', while for the lateral tendon part LWk was low, corresponding to only 'fair' agreement'.

Strain Ratios and Raw data

The ROI's with raw SEL data (no reference tissue) resulted in the highest reliability in intra- as well as inter-rater reliability. Although there was a significant difference between Rater 1 and Rater 2 for RAW, this difference was less than the MDC and is ascribed as measurement error.

To our knowledge no study has previously presented raw SEL data, which hampers comparison with other studies. Using raw data has limitations, since there is, in comparison with using ratios, no possibilities to adjust for different transducer pressures. On the other hand, the advantage is that it gives a more quantitative estimate than when using visual inspections, as for example in the color scale, or uncertainties from selection of reference area.

In the present study only one rater captured the images (using almost same pressure) and two raters traced and measured the areas. If more than one rater were to capture the images a lower reliability may be expected, due to potentially different transducer pressure.

Different reference tissues in SEL measurements of the musculoskeletal area make it difficult to compare data across studies. The reference tissues previously used for calculating strain ratios in the achilles- and supraspinatus tendon areas have included bone,[15], fat,[13, 25, 31, 42, 43] and gel pad/acoustic coupler,[13, 16, 44]. Using bone as reference value has limitations, since ultrasound cannot penetrate bone material, meaning that data coming from this region is artefacts.

In addition, strain ratios may in different studies be based on different equations, placing the ROIs of the tendon in the denominator,[13, 45] or, as in the present study, in the numerator,[25, 31].

Due to thin subcutaneous fat tissue area in some participants the reference tissue was replaced, after phase one, into muscle tissue (deltoid muscle) and the gel pad (artificial fat tissue).

When using the deltoid muscle tissue (DELT) as reference, reliability was found to be excellent (both intra- and inter-rater). Muscle tissue increases stiffness significantly after exercise and muscle contractions,[16, 24] and therefore limitations are recognized, when using this tissue as reference for investigating e.g. tendon tissue response to muscle contractions.

However no previous study of SEL on the supraspinatus tendon has used muscle tissue as reference for the calculated strain ratios.

The present study found a high MDC for GEL (large measurement error) but a good reliability (ICC 0.70-0.73), in line with previous studies. One study, using the Kager's fat pad (fat deposit within the Kager triangle anterior to the achilles tendon), as reference tissue in the achilles tendon, also found good to excellent intra- and inter-tester reliability (ICC: 0.51-0.78) in the longitudinal plane, however, with much lower reliability (ICC: 0.41-0.45) in the axial plane,[25]. Unfortunately MDC was not reported.

The present data with the gel pad are also in line with a study using an acoustic coupler as reference tissue in the supraspinatus tendon,[16], where excellent intra-rater reliability was shown, as also confirmed for the achilles tendon,[44]. An acoustic coupler is similar to the gel pad and may be acceptable, but ideally, the reference area should be in the same depth as the ROI (here the supraspinatus tendon). As the gel pad is not located closely to the tendon, the ROIs from the GEL pad will not be subjected to the same amount of tissue pressure as the tendon, which may affect reliability and validity.

The present study found the lowest reliability (but still graded 'good') when using GEL which may be caused by difficulties due to lower image quality, because of increased depth. Also a statistical significant intra-rater difference was seen, but as this difference was below the MDC it can be ascribed as measurement error.

Color Ratings

The present study found a high reliability ('substantiel' to 'almost perfect') when grading the ROIs of SEL images according to a 4-level color scale where blue tissue indicates hard tissue and red tissue indicates soft tissue. This is in line with a previous study on the achilles tendon using color scales of 5-levels (1=blue (hardest tissue), 2=light blue, 3=green, 4=yellow, 5=red

(softest tissue)), where good to excellent inter-rater (κ = 0.897) and intra-rater (κ = 1) reliability was found,[30].

Alternative types of color scales have been used in the musculoskeletal area, primarily for the achilles

tendon, with categorization of the achilles tendon into a 2-level category scale (green/blue vs. red),[25] and a 3-level category (blue/yellow/red),[22, 29]. The current study has used the same colors for defining hard tissue (blue), and for soft tissue (red), but with a 4-level category scale, and with more precise criteria for the different levels which was found necessary due to the mixture of colors in the supraspinatus tendon observed in phase 1.

In the present study LWk in the medial and the middle tendon part was 'substantial to almost perfect' when counting the number of yellow/red lesions in both intra and inter-rater reliability. Even though the total agreement was high also in the lateral third of the tendon, LWk was relatively low compared to the medial and the middle part of the tendon. The reason may be due to the low presence of lesions in the lateral part, corresponding to only 10 % of the participants presenting with lesions in the lateral 1/3 of the tendon which can lead to the 'Kappa Paradox',[46]. Another explanation is that since the pressure is put vertically on the medial part of the tendon, the lateral part will have experienced a smaller degree of stress. The high reliability of number of focal lesions (LwK intra-rater reliability: 87.5-90, LwK inter-rater reliability: 82.5-90) is in line with a previous study on the supraspinatus tendon where an almost prefect reliability (k = 0.83) was found,[14].

The present methods of using color grading and lesion counts for assessing tendon stiffness in the supraspinatus tendon showed high reliability. These methods are feasible for use in clinical practice as they can be performed quickly, but when more than one clinician is performing the SEL, the method can not adjust for potentially different tranducer pressures.

Limitations

Manual compression may affect reliability, especially when using raw data, color grading and counting yellow/red lesions. To partly counteract for this, a quality bar was used providing instant feedback on the uniformity of the transducer pressure. Further, also reference tissues (deltoid muscle and gel pad) with calculation of strain ratios were used, thereby making a comparison possible between the methods.

Furthermore, as mentioned, SEL is highly operator dependent, why the same (and only one) trained operator captured the present images, thereby further decreasing risk of bias, but this limits the external validity of the results.

In addition, age was not blinded for and as earlier reported; tendons get softer with age,[47]. Therefore the current blinding of examiner results and health status may not have been sufficient.

Strengths

The strength of this study is the design, incorporating a stepwise and standardized procedure for reliability which minimizes bias and increases reliability,[18]. Furthermore both patients and healthy participants were enrolled.

To enhance standardization all SEL ROIs were measured at a fixed point just laterally from the anterior-lateral corner of the acromion in the longitudinal plane of the supraspinatus

tendon [20]. In addition the reliability was estimated by using three different quantitative methods (RAW, DELT and GEL), as well as two different qualitative methods (color and number of yellow/red focal lesions).

CONCLUSION

Intra- and inter-rater reliability were excellent using raw values and ratios with deltoid as reference tissue, and good when using gel pad as reference.

Reliability of color scale ratings, were substantial to almost perfect and number of lesions were fair to almost perfect.

Although high reliability was found, validity and responsiveness of these elastographic methods must be established.

Contributorship statement
KB, JH, PK, KGI and BJ-K conceived and designed the study protocol. KB and BJ-K procured the project funding. KB, JH, PK, and BJ-K developed and standardised the ultrasound procedure. JH, KB and KGI recruited participants.
KB was the project coordinator and captured the strain elastography images. KGI performed the physical tests. KB and JH rated the images. KB and BJ-K planned and coordinated the statistical analyses. KB performed the statistical analyses. KB drafted the manuscript, and JH, PK, KGI and BJ-K contributed to the manuscript. All authors read and approved the final manuscript. KB is the guarantor.

Competing interests
The authors declare no conflicts of interest.

Funding
This work was supported by South Denmark Region (14/38969, 15/24799), The Danish Rheumatism Association (R150-A4296) and Danish Council of Radiographers (no grant number).

Data sharing statement
No additional data is available.

REFERENCES

1. Luime JJ, Koes BW, Hendriksen IJ, Burdorf A, Verhagen AP, Miedema HS et al. Prevalence and incidence of shoulder pain in the general population; a systematic review. Scand J Rheumatol 2004;33:73-81.

2. Singh SG, Sumit.; Mohammad, F.; Kumar, S.; Kumar, D.; Kumar, S. Prevalence of shoulder disorders in tertiary care centre. J Res Med Sci 2015;3:917-920.

3. Nyman PP, K.; Panula, H.; Mälkiä, E. Patients' experience of shoulder problems prior to and following intervention. Physiother Theory Pract 2012;28:221-231.

4. Virta L, Joranger P, Brox JI, Eriksson R. Costs of shoulder pain and resource use in primary health care: a cost-of-illness study in Sweden. BMC Musculoskelet Disord 2012;13:17

5. Teunis T, Lubberts B, Reilly BT, Ring D. A systematic review and pooled analysis of the prevalence of rotator cuff disease with increasing age. J Shoulder Elbow Surg / A2014;23:1913-1921.

6. Klauser AS, Faschingbauer R, Jaschke WR. Is sonoelastography of value in assessing tendons? Semin Musculoskelet Radiol 2010;14:323-333.

7. Prado-Costa R, Rebelo J, Monteiro-Barroso J, Preto AS. Ultrasound elastography: compression elastography and shear-wave elastography in the assessment of tendon injury. Insights imaging 2018.

8. Dietrich CF, Barr RG, Farrokh A, Dighe M, Hocke M, Jenssen C et al. Strain Elastography - How To Do It? Ultrasound Int Open 2017;3:E137-e149.

9. Garra BS. Elastography: history, principles, and technique comparison. Abdom Imaging 2015;40:680-697.

10. Chung SW, Kim JY, Yoon JP, Lyu SH, Rhee SM, Oh SB. Arthroscopic repair of partial-thickness and small full-thickness rotator cuff tears: tendon quality as a prognostic factor for repair integrity. Am J Sports Med 2015;43:588-596.

11. Bamber J, Cosgrove D, Dietrich CF, Fromageau J, Bojunga J, Calliada F et al. EFSUMB guidelines and recommendations on the clinical use of ultrasound elastography. Part 1: Basic principles and technology. Ultraschall Med 2013;34:169-184.

12. Fusini F, Langella F, Busilacchi A, Tudisco C, Gigante A, Masse A et al. Real-time sonoelastography: principles and clinical applications in tendon disorders. A systematic review. Muscles Ligaments Tendons J 2017;7:467-477.

13. Lee SU, Joo SY, Kim SK, Lee SH, Park SR, Jeong C. Real-time sonoelastography in the diagnosis of rotator cuff tendinopathy. J Shoulder Elbow Surg 2016;25:723-729.

14. Seo JB, Yoo JS, Ryu JW. Sonoelastography findings of supraspinatus tendon in rotator cuff tendinopathy without tear: comparison with magnetic resonance images and conventional ultrasonography. J Ultrasound 2015;18:143-149.

15. Tudisco C, Bisicchia S, Stefanini M, Antonicoli M, Masala S, Simonetti G. Tendon quality in small unilateral supraspinatus tendon tears. Real-time sonoelastography correlates with clinical findings. Knee Surg Sports Traumatol Arthrosc 2013;23(2):393-8.

16. Muraki T, Ishikawa H, Morise S, Yamamoto N, Sano H, Itoi E et al. Ultrasound elastography-based assessment of the elasticity of the supraspinatus muscle and tendon during muscle contraction. J Shoulder Elbow Surg 2015;24:120-126.

17. Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hrobjartsson A et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. J Clin Epidemiol 2011;64:96-106.

18. Patijn JV, Beek JV, Blomberg S, et al. Reproducibility and Validity Studies of Diagnostic Procedures in Manual/Musculoskeletal Medicine; 2004. In: Patijn J, ed. International Federation for manual/musculoskeletal medicine. 2004:3–35.

19. Hougs Kjaer BPM, Ellegaard KPP, Wieland IP, Warming SPP, Juul-Kristensen BPP. Intra-rater and inter-rater reliability of the standardized ultrasound protocol for assessing subacromial structures. Physiother Theory Pract 2017;33:398-409.

20. Ingwersen KG, Hjarbæk J, Eshøj H, Larsen CM, Vobbe JW, Juul-Kristensen B. Ultrasound assessment for grading structural tendon changes in supraspinatus tendinopathy: an inter-rater reliability study. B M J Open 2016;6. 10.1136/bmjopen-2016-011746.

21. Sein ML, Walton J, Linklater J, Harris C, Dugal T, Appleyard R et al. Reliability of MRI assessment of supraspinatus tendinopathy. "Br J Sports Med 2007;41:e9. doi:10.1136/bjsm.2006.034421

22. De Zordo T, Fink C, Feuchtner GM, Smekal V, Reindl M, Klauser AS. Real-time sonoelastography findings in healthy Achilles tendons. AJR 2009;193:W134-138.

23. Cosgrove D, Piscaglia F, Bamber J, Bojunga J, Correas JM, Gilja OH et al. EFSUMB guidelines and recommendations on the clinical use of ultrasound elastography. Part 2: Clinical applications. Ultraschall Med 2013;34:238-253.

24. Niitsu M, Michizaki A, Endo A, Takei H, Yanagisawa O. Muscle hardness measurement by using ultrasound elastography: a feasibility study. Acta Radiol 211;52:99-105.

25. Drakonaki EE, Allen GM, Wilson DJ. Real-time ultrasound elastography of the normal Achilles tendon: reproducibility and pattern description. Clin Radiol 2009;64:1196-1202.

26. Lumsdaine W, Smith A, Walker RG, Benz D, Mohammed KD, Stewart F. Morphology of the humeral insertion of the supraspinatus and infraspinatus tendons: Application to rotator cuff repair. Clin Anat 2015;28:767-773.

27. Curtis AS, Burbank KM, Tierney JJ, Scheller AD, Curran AR. The insertional footprint of the rotator cuff: an anatomic study. Arthroscopy 2006;22:609.e601.

28. De Zordo T, Lill SR, Fink C, Feuchtner GM, Jaschke W, Bellmann-Weiler R et al. Real-time sonoelastography of lateral epicondylitis: comparison of findings between patients and healthy volunteers. AJR 2009;193:180-185.

29. Klauser AS, Miyamoto H, Tamegger M, Faschingbauer R, Moriggl B, Klima G et al. Achilles tendon assessed with sonoelastography: histologic agreement. Radiology 2013;267:837-842.

30. Sconfienza LM, Silvestri E, Cimmino MA. Sonoelastography in the evaluation of painful Achilles tendon in amateur athletes. Clin. Exp. Rheumatol 2010;28:373-378.

31. Turan A, Tufan A, Mercan R, Teber MA, Tezcan ME, Bitik B et al. Real-time sonoelastography of Achilles tendon in patients with ankylosing spondylitis. Skeletal Radiol 2013;42:1113-1118.

32. Kitis A, Celik E, Aslan UB, Zencir M. DASH questionnaire for the analysis of musculoskeletal symptoms in industry workers: a validity and reliability study. Appl Ergon 2009;40:251-255.

33. Boonstra AM, Schiphorst Preuper HR, Reneman MF, Posthumus JB, Stewart RE. Reliability and validity of the visual analogue scale for disability in patients with chronic musculoskeletal pain. Int J Rehabil Res 2008;31:165-169.

34. Norman R, Cronin P, Viney R, King M, Street D, Ratcliffe J. International comparisons in valuing EQ-5D health states: a review and analysis. Value Health 2009;12:1194-1200.

35. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86:420-428.

36. Cicchetti D. Guidelines, Criteria, and Rules of Thumb for Evaluating Normed and Standardized Assessment Instrument in Psychology. Psychol Assess 1994;6:284-290.

37. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307-310.

38. Ottenbacher KJ, Stull GA. The analysis and interpretation of method comparison studies in rehabilitation research. Am J Phys Med Rehabil 1993;72:266-271.

39. de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. J Clin Epidemiol 2006;59:1033-1039.

40. deVet HCW, Terwee, Caroline B.; Knol, Dirk L.;, Mokkink, Lidwine B. . Measurements in Medicine - a pratical guide. New York, USA: Cambridge University Press; 2011.

41. Landis RJK, G.G. The Measurement of Observer Agreement for Categorical Data. Biometrics 1977;33:159–174.

42. Busilacchi A, Olivieri M, Ulisse S, Gesuita R, Skrami E, Lording T et al. Real-time sonoelastography as novel follow-up method in Achilles tendon surgery. Knee Surg Sports Traumatol Arthrosc 2014.

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

44. Yamamoto Y, Yamaguchi S, Sasho T, Fukawa T, Akatsu Y, Nagashima K et al. Quantitative Ultrasound Elastography With an Acoustic Coupler for Achilles Tendon Elasticity: Measurement Repeatability and Normative Values. J Ultrasound Med 2016;35:159-166.

45. Akkaya S, Akkaya N, Agladioglu K, Gungor HR, Ok N, Ozcakar L. Real-time elastography of patellar tendon in patients with auto-graft bone-tendon-bone anterior cruciate ligament reconstruction. Arch Orthop Trauma Surg 2016;136:837-842.

46. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol 1990;43:543-549.

47. Hsiao MY, Chen YC, Lin CY, Chen WS, Wang TG. Reduced patellar tendon elasticity with aging: In vivo assessment by shear wave elastography. Ultrasound Med Biol 2015;41:2899-2905.

Figure legends

Figure 1

With the arm located behind the back and the elbow flexed 90° with the palm facing towards the posterior direction, the probe is placed just laterally from the anterior-lateral corner of the acromion in the longitudinal plane of the supraspinatus tendon.

Figure 2

Left side: For measuring elastography characteristics of the supraspinatus tendon, the tendon was split into 3 parts (3x7.7mm), illustrated by the areas of blue, red and green colors. This division was based on a line from the lateral tendon insertion (tuberculum majus) to the medial tendon insertion part, corresponding to 6.5 mm, and from there a line of 23 mm to the medial tendon part with the end point (medial part) being perpendicular to the superior surface of the tendon.

The yellow circle in the soft part of the deltoid muscle is used as a reference.

Right side: The three measurements areas (ROI's) and one reference area with elastography characteristics (raw data) during the time of measurement.

Figure 3

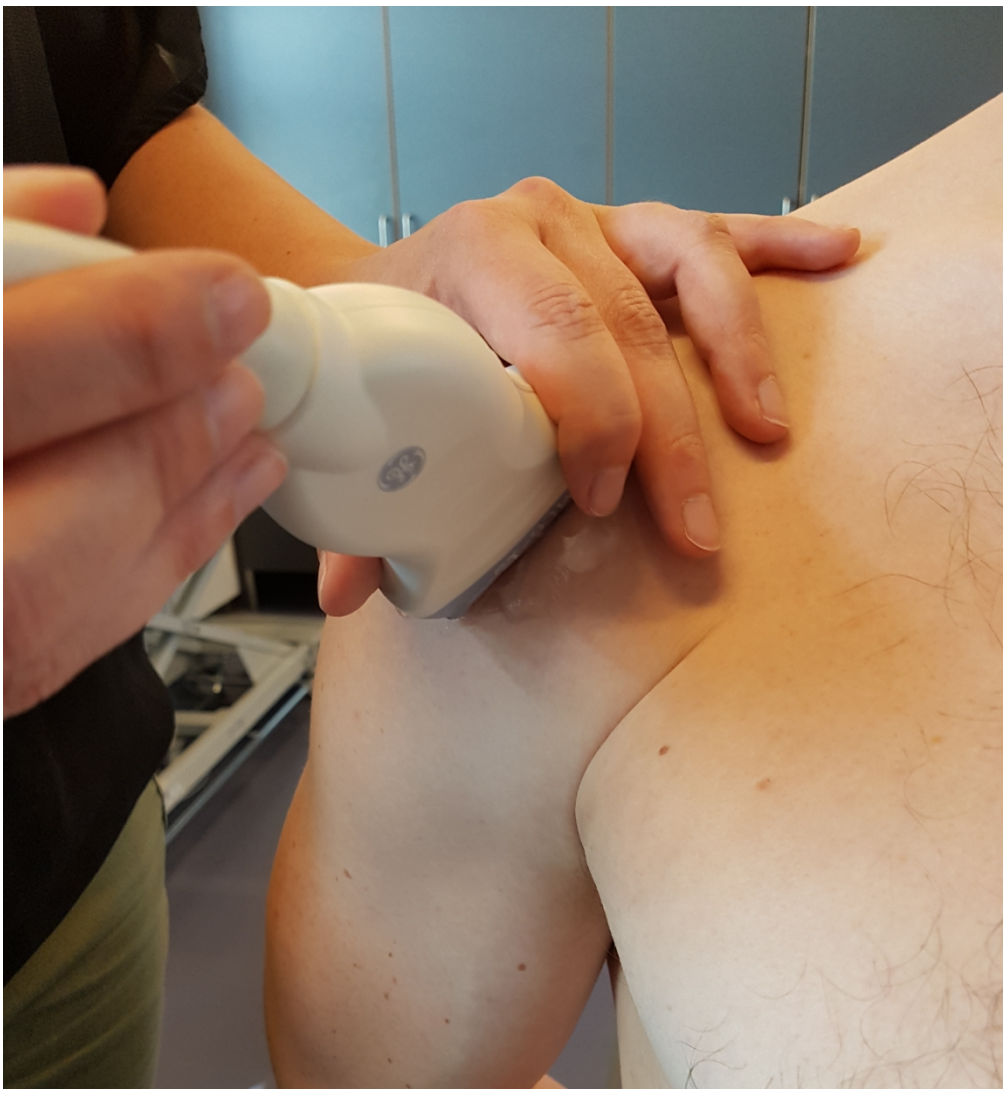
Left side: Gel pad (mounted on the transducer with a condom used as the other reference area, yellow circle). Right side: The corresponding elastography image/measurements.

Figure 4

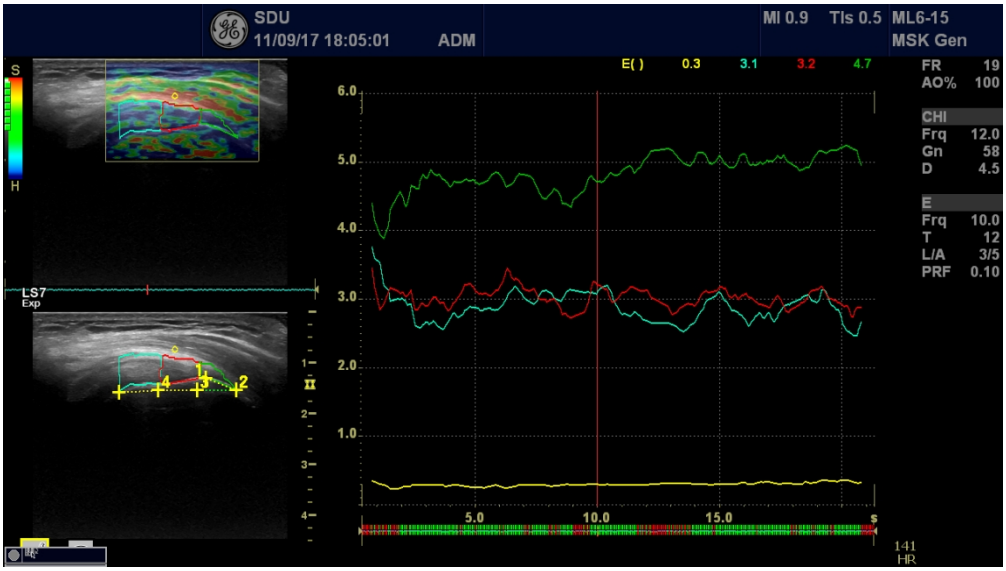
A 6.5 mm chord was drawn (in the lateral direction) from the medial part of the insertion on the humeral head to the lateral part of the tuberculum major. This fix point (end of the line at the lateral part of the insertion) was used to draw a 23 mm (7.7 x 3) horizontal line in the medial direction ensuring agreement of measuring area.

Figure 5

Bland-Altman plots with 95% Limits of Agreement for measurements, using respectively raw data, the deltoid muscle and a gel pad as reference areas. Values are based on the mean of the three measured parts across the supraspinatus tendon.

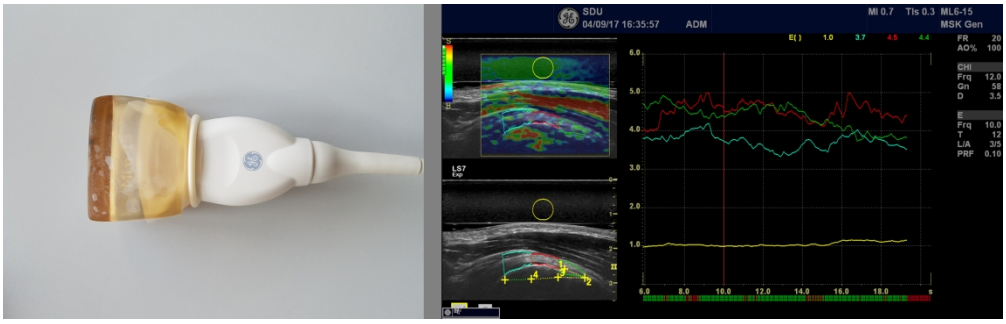


With the arm located behind the back and the elbow flexed 90° with the palm facing towards the posterior direction, the probe is placed just laterally from the anterior-lateral corner of the acromion in the longitudinal plane of the supraspinatus tendon.

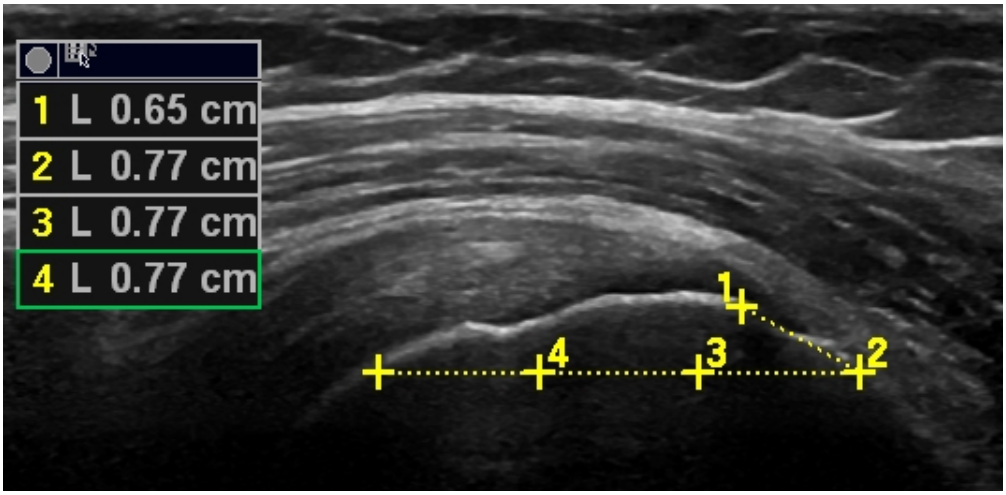


Left side: For measuring elastography characteristics of the supraspinatus tendon, the tendon was split into 3 parts (3x7.7mm), illustrated by the areas of blue, red and green colors. This division was based on a line from the lateral tendon insertion (tuberculum majus) to the medial tendon insertion part, corresponding to 6.5 mm, and from there a line of 23 mm to the medial tendon part with the end point (medial part) being perpendicular to the superior surface of the tendon.
The yellow circle in the soft part of the deltoid muscle is used as a reference.
Right side: The three measurements areas (ROI's) and one reference area with elastography characteristics (raw data) during the time of measurement.

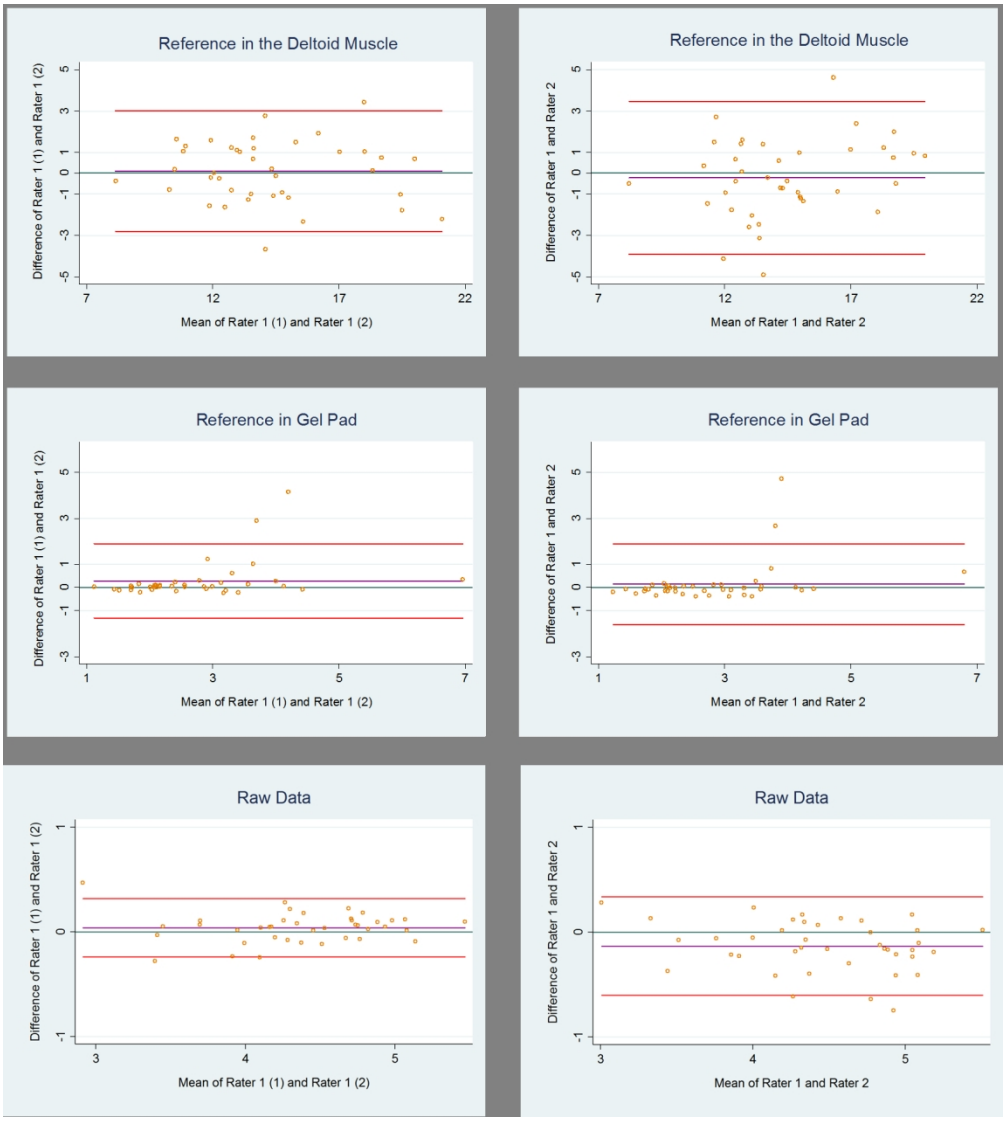
451x254mm (72 x 72 DPI)



Left side: Gel pad (mounted on the transducer with a condom used as the other reference area, yellow circle). Right side: The corresponding elastography image/measurements.



A 6.5 mm chord was drawn (in the lateral direction) from the medial part of the insertion on the humeral head to the lateral part of the tuberculum major. This fix point (end of the line at the lateral part of the insertion) was used to draw a 23 mm (7.7 x 3) horizontal line in the medial direction ensuring agreement of measuring area.



Bland-Altman plots with 95% Limits of Agreement for measurements, using respectively raw data, the deltoid muscle and a gel pad as reference areas. Values are based on the mean of the three measured parts across the supraspinatus tendon.

BMJ Open

ULTRASONIC STRAIN ELASTOGRAPHY FOR DETECTING ABNORMALITIES IN THE SUPRASPINATUS TENDON: AN INTRA- AND INTER-RATER RELIABILITY STUDY

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-027725.R1
Article Type:	Research
Date Submitted by the Author:	28-Feb-2019
Complete List of Authors:	Brage, K; University of Southern Denmark, Sport Science and Clinical Biomechanics Hjarbaek, John; Odense Universitetshospital, Department of Radiology, Musculoskeletal section Kjaer, Per ; University of Southern Denmark, Institute of Sports Science and Clinical Biomechanics Ingwersen, Kim G.; Sports Science and Clinical Biomechanics Juul-Kristensen, Birgit; Research Unit for Musculoskeletal Function and Physiotherapy, Institute of Sports Science and Clinical Biomechanics
Primary Subject Heading:	Radiology and imaging
Secondary Subject Heading:	Diagnostics
Keywords:	Strain elastography, supraspinatus tendon, rotator cuff, tendon quality, reliability, tendinopathy

ULTRASONIC STRAIN ELASTOGRAPHY FOR DETECTING ABNORMALITIES IN THE SUPRASPINATUS TENDON: AN INTRA- AND INTER-RATER RELIABILITY STUDY

Brage K[1,2], Hjarbaek J[3], Kjaer P[1,2], Ingwersen KG[4], Juul-Kristensen B[1]

[1] Department of Sports Science and Clinical Biomechanics, University of Southern Denmark, Campusvej 55, 5230 Odense M, DK
[2] UCL University College, Niels Bohrs Allé 1, 5230, Odense M, DK
[3] Department of Radiology, Odense University Hospital, J. B. Winsløws Vej 4, 5000 Odense C, DK
[4]Research Unit in Physiotherapy and Occupational Therapy, Hospital Lillebaelt, Kabbeltoft 25, 7100 Vejle, DK

Author information
Corresponding Author
Karen Brage, PhD candidate, Telephone: +45 51787381, fax: 6318 3226, e-mail: kabr2@ucl.dk
Co-Authors
John Hjarbaek, MD, Telephone: +45 66113333, e-mail: john.hjarbaek@rsyd.dk
Per Kjaer, Professor, PhD., Telephone: +45 65504553, e-mail: pkjaer@health.sdu.dk
Kim Gordon Ingwersen, PhD., Telephone: +45 65509044, e-mail: kingwersen@health.sdu.dk
Birgit Juul-Kristensen, PhD.: Telephone: +45 65503412, e-mail: bjuul-kristensen@health.sdu.dk

Wordcount
4404

Trial registration
The study protocol was approved by the Ethics Committee for the Region of South Denmark (S-20160115) and reported to the Danish Data Protection Agency (2014-41-3266).

ABSTRACT

Objectives

The reliability of ultrasonic strain elastography (SEL) used to detect abnormalities in the supraspinatus tendon is unclear. Thus, the aim of this study was to investigate the reliability of SEL in the supraspinatus tendon.

Design

An intra-rater and inter-rater reliability study.

Setting

A single-centre study conducted at the University of Southern Denmark.

Participants

Twenty participants with shoulder pain and MRI-verified supraspinatus tendinosis and 20 asymptomatic participants (no MRI).

Primary and secondary outcome measures

Raw values (RAW), and ratios (deltoid muscle (DELT); gel pad (GEL) as reference tissues) were calculated and mean values of measurements from three regions of the supraspinatus tendon were reported. Colour scale ratings and number of yellow/red lesions from the three areas were also included.

Results

Intra-rater reliability showed Intra-Class Correlation Coefficients (ICCs) for RAW, DELT and GEL: 0.97 (Minimal Detectable Change (MDC): 0.28 (6.36% of the mean); 0.89 (MDC: 2.91 (20.37%); and 0.73 (MDC: 1.61 (58.82%)), respectively. The ICCs for inter-rater reliability were 0.89 (MDC: 0.47 (10.53%); 0.78 (MDC: 3.69 (25.51%); and 0.70 (MDC: 1.75 (62.63%)), respectively.

For colour scale ratings, intra-rater reliability (Linear Weighted kappa, LWk) ranged from 0.76 to 0.79, with the inter-rater reliability from 0.71 to 0.81. For the number of lesions, intra-rater reliability ranged from 0.40 to 0.82, and inter-rater reliability from 0.24 to 0.67.

Conclusions

Intra-rater and inter-rater reliability were excellent for raw values and for ratios with deltoid muscle as the reference tissue, and good for ratios with gel pad as the reference tissue. The reliability of colour scale ratings was substantial to almost perfect, and for the number of lesions fair to almost perfect.

Although high reliability was found, validity and responsiveness of these elastographic methods needs further investigation.

Strengths and limitations of this study

- Standardised procedure for capturing strain elastographic images was developed and applied
- Raw values of strain elastographic images were presented
- Ratios, based on different reference tissues or areas (deltoid muscle/gel pad), of strain elastographic images were calculated and presented
- Specific procedures for colour grading strain elastographic images were presented
- SEL is highly operator-dependent

INTRODUCTION

Shoulder pain is a common symptom, with a lifetime prevalence in the general population of 6.7-66.7%[1] and subacromial pain syndrome is the second most common cause of pain in the shoulder.[2] Shoulder pain has consequences such as physical limitations, mental problems[3] and absence from work.[4]

Shoulder disorders are evaluated by history-taking and physical examination, potentially supplemented with X-ray, conventional ultrasound and/or Magnetic Resonance Imaging (MRI). However, supraspinatus tendon abnormalities are also found in asymptomatic people when using general modalities such as MRI and ultrasound.[5] Furthermore, it seems difficult to distinguish pathological changes from healthy tissues by using conventional ultrasound because pathological regions often exhibit the same echogenicity as non-pathological regions.[6]

Strain elastography (SEL) is a relatively new and not yet well-established method, which may assist in diagnosis, prediction and monitoring of progress in tendon healing.[7]

SEL defines the physical properties of soft tissues through characterisation of the differences in stiffness between 'the regions of interest' (ROI) and the surrounding tissues.[8 9] Conventional ultrasound and MRI were developed for visualisation of macroscopic changes and not specifically for the mechanical tendon properties, which is why SEL may add further knowledge to conventional shoulder imaging. Since tendon quality is a prognostic factor for rotator cuff repair, information about tendon stiffness could be beneficial for the surgeon.[10]

Tissue deformation using SEL is obtained by uniform mechanically induced compressions (strain) of the structures under the ultrasound transducer, during the ultrasound scan. Through manual

compression, the soft tissue deforms differently depending on its inherent stiffness. The degree of deformation can be interpreted as an estimate of the tissue stiffness.[11]

SEL has recently been used in the musculoskeletal area, where the achilles tendon has been the primary focus.[12] A few studies have found significant associations between pathology identified by SEL and MRI in patients with supraspinatus tendinopathy,[13 14] besides significant correlations between results from SEL and clinical tests and questionnaires in patients with small supraspinatus tendon tears.[15] One study also found SEL to be able to detect increased stiffness in the supraspinatus tendon elasticity and muscle elasticity with increased muscle contraction in healthy participants.[16] Although only a few studies have been performed, and with different aims, comparator instruments, procedures, reference tissues and data types, the concurrent validity of SEL in the supraspinatus tendon seems promising.[13-16]

However, the reliability of SEL must first be proven to be acceptable. In this respect, SEL constitutes several challenges since SEL is a technique with relatively high operator dependency in terms of the manually applied pressure and subsequent identification and selection of the pathological ROI.

One of only two reliability studies found the inter-rater reliability of SEL in the supraspinatus tendon to be almost perfect (κ = 0.83) with respect to the number of focal lesions in 118 patients with MRI-verified supraspinatus tendinopathy.[14] However, this study did not include a healthy control group which is recommended for reliability and validity studies, to achieve realistic tissue variation.[17] Furthermore, the study used only colour quantification.[14] The other study found a high intra-rater reliability (ICC $_{1,3}$ = 0.92-0.99) with respect to ROI, when using an acoustic coupler as

the reference tissue in a small sample of 23 healthy participants. The limitations of this study were: not including a group with pathology, not defining ROI and only using one type of reference tissue (acoustic coupler).[16] Since standardised and consensus procedures for conducting elastography in the supraspinatus tendon have not yet been established, there is a need to investigate which reference tissue has the highest reliability. To our knowledge, there has not been a study investigating the reliability of SEL in the supraspinatus tendon that has included both patients with a pathological (non-ruptured) supraspinatus tendon and healthy participants with non-painful shoulders. Furthermore, choice of reference tissue and quantification methods has major impact on results, but no reliability studies have compared these different approaches.

The aim of this study was to test the intra-rater and inter-rater reliability of SEL within the supraspinatus tendon in patients and healthy participants, using different reference tissues (deltoid muscle (DELT), gel pad (GEL)), and different quantification methods (raw data, strain ratios, colour scale rating and counting number of yellow/red lesions).

MATERIALS AND METHODS

This study was an intra-rater and inter-rater reliability study of SEL, used on the supraspinatus tendon, reported according to the Guidelines for Reporting Reliability and Agreement Studies (GRRAS).[18]

A three-phased intra-rater and inter-rater reliability protocol for diagnostic reliability studies was used.[17] The protocol included a training phase (Phase One), an overall agreement phase (Phase Two), and an actual study phase (Phase Three), for securing low clinician dependency and subjectivity, sufficient experience and standardisation, and minimisation of systematic bias.[17] In order to progress to the study phase, the criterion of at least 80% inter-rater agreement in Phase Two was used. Such a protocol has previously been used in reliability studies, using ultrasound methods.[19 20]

Study procedures

The study's participants were recruited from August 2016 to December 2017. Symptomatic participants, with an MRI-verified supraspinatus tendinosis (*patients)*, were recruited from the Radiology Department at Odense University Hospital within 14 days of their MRI examination. Participants with no shoulder symptoms (*healthy participants)* were recruited primarily from the University of Southern Denmark and the UCL University College, both situated in Odense. Except for MRI, all the procedures were performed at the University of Southern Denmark. After inclusion, participants underwent clinical tests (performed by KGI) and filled out questionnaires regarding functional limitations, pain and quality of life. Testing procedures took place once and lasted approximately one hour.

All SEL images were captured by the same clinician (KB), while two raters performed the SEL measurements on the captured images. Rater 1 (KB), radiographer, was thoroughly trained in using SEL on the shoulder by Rater 2 (JH), radiologist, who had more than 20 years' experience in clinical musculoskeletal ultrasound. In Phase Three, the raters were blinded to each other's results (data were stored separately), the participants' health status (Rater 1 entered the room after clinical tests and questionnaires were performed, and Rater 2 had no contact with the participants) and their MRI results (ID-numbers were changed after MRI examinations).

SEL images were stored for at least 14 days after image capturing until the first image assessment, to ensure elimination of any memory of pain response during SEL by Rater 1. Further, all SEL images were stored for an additional 14 days before Rater 1's second assessment, to ensure elimination of recalls of SEL results from the first assessment.

The study protocol was approved by the Ethics Committee for the Region of South Denmark (S-20160115) and reported to the Danish Data Protection Agency (2014-41-3266). All participants had oral and written information about the study and signed an informed consent form before participation in this study.

Patient and Public Involvement

Patients and public were not involved in this study.

Participants

Inclusion and exclusion criteria

An inclusion criterion for both patients and healthy participants was age 18-65 years. In addition, patients had to have at least one shoulder clinically diagnosed with tendinosis (positive signs from ≥ 3 clinical tests (Hawkins-Kennedy test; Neers test; Empty Can test; Full Can test; Resisted external rotation test)) and based on MRI (≥ Grade 1; corresponding to focal increase in tendon signal on protondensity-weighted and fat-suppressed T2 sequences not equal to that of fluid.[21] If both shoulders were MRI-scanned, the most severely affected side was included.

Inclusion criteria for the healthy participants were no previously experienced shoulder problems, and negative signs from all five clinical tests described above. For these participants, the choice of shoulder was matched with that of the patients.

Exclusion criteria for both groups were: Tears >1/3 of the vertical height of the supraspinatus tendon, since the stress may be increased on the intact tendon part, and calcifications >2 mm (length) due to acoustic shadowing. Further exclusion criteria were: previous comorbidities (potentially harmful to the tendon) such as; past/present shoulder fractures, surgery and luxation, known neuromuscular disease, rheumatoid arthritis, cancer, fibromyalgia, spondyloarthropathy and psychiatric disorders. Pregnancy and inability to read and understand Danish were also exclusion criteria.

SEL image capturing and measurement

Apparatus

All measurements were performed with the Logiq S7 using a 15 MHz linear probe (GE Healthcare, Milwaukee, USA). Manufacturer recommendations for musculoskeletal SEL of the shoulder were used, including a transducer movement rate of approximately 120 cycles/min, axial smoothing of 2, lateral smoothing of 3, frequency of 10 and a soft/hard compression of 5.

Patient placement

The SEL was obtained with the patient sitting in an erect posture with the arm internally rotated, elbow flexed to 90° and with the dorsal side of the hand placed over the sacrum, as previously used.[15] The probe was placed on the anterior aspect of the acromion in the coronal plane and the images were obtained just laterally to the anterior-lateral corner of the acromion in the longitudinal plane of the supraspinatus tendon (Figure 1).

Image capturing

An image window depth of at least three times the tendon size and an image width covering about three-quarters of the screen were used as recommended for longitudinal SEL.[22] The tissue was compressed approximately 2-5 mm,[9] and a software incorporated quality control (expressed as 1-5 green bars being displayed, with 5 bars being the most acceptable) was used to evaluate the recommended compression size.

For each assessment method (with or without GEL covering the transducer, (Sonokit (Proxon), thickness: 10 mm, length: 70 mm, elastic modulus: 226 kPa; Sonogel Vertriebs GmbH, Germany)), three sessions of 20 seconds were obtained.

Image measurements

Tendon characteristics were evaluated quantitatively and qualitatively.

Quantitatively, ROIs on the SEL images were drawn over the target area (supraspinatus tendon) and the exact raw strain value (RAW) (0-6; 6 being the hardest tissue) was calculated.

Further, an adjacent reference region (normal tissue, experiencing the same stress as the target region) was drawn. From these two variables, the strain ratio (0-60; 60 being the hardest tissue) was calculated.[23] Two different reference tissues were used: a 1 mm circle region in a soft part of the deltoid/bursal area for the DELT measurement (Figure 2), and a 5 mm circle region in GEL covering the transducer, for the GEL measurement (Figure 3).[13 24] GEL was used as a more homogeneous reference tissue.

For the raw values alone and the strain ratios (strain raw value of supraspinatus tendon (A) to strain raw value of reference tissue (B) = A/B), a mean of the three measured areas of the supraspinatus tendon was calculated including data from 5-15 seconds of the 20-second cycle as recommended by the manufacturer.

Quantitative measurements were based on examination of three entire cine-loops (10 seconds) rather than on single static images,[25] in order to minimise intra-observer variation and avoid transient temporal fluctuations. Only sequences with the highest image quality (with green bars on the quality assessment) were used as recommended by the manufacturer.

Due to difficulties in defining the most lateral part of the supraspinatus insertion on the humeral head, a 6.5 mm chord was drawn (in the lateral direction) from the medial part of the insertion on the humeral head to the lateral part of the tuberculum major. The average length of the supraspinatus insertion has previously been estimated to be 6.5 mm.[26] This fix point (end of the line at the lateral part of the insertion) was used to draw a 23 mm (7.7 x 3) horizontal line in the medial direction (which has been estimated to be the average length of the tendon),[27] ensuring agreement of measurement area (Figure 4).

Caudal borders for image measurements were bony surfaces (of the humeral head), while cranial borders were the transition zone between the superior surface of the supraspinatus tendon and the inferior surface of the deltoid muscle and bursa. The bursa area was used as reference tissue (red colour in the SEL image). If no red border was seen, an estimated border followed the superior surface of the tendon.

Qualitatively, the images were rated using a colour scale, from 1-4, according to the following: Type 1, < 10% colour other than blue (indicating a high tissue stiffness); Type 2, 11-25% colour other than blue; Type 3, 26-50% colour other than blue; or Type 4, > 50% colour other than blue (indicating a low tissue stiffness). Furthermore, the number of yellow/red lesions was graded as follows: 0=no lesions; 1=one lesion; 2=two lesions; and 3=more than two lesions.[14] The qualitative classifications were all performed on the first high quality image recorded closest to 10 seconds into the first cycle.

Development of SEL method

Based on previous studies,[14 16 25 28-31] a protocol with standardised procedures was developed for obtaining SEL characteristics in the Supraspinatus tendon and tested in Phase One on 10 participants.

Based on the results from these participants, adjustments to the type of reference tissue and colour scale criteria were performed. One adjustment involved the replacement of the subcutaneous fat with DELT as reference tissue, since the subcutaneous fat layer in some participants was too thin to measure. Another adjustment was to base the colour scale on the

percentage of colour (replacement of estimation of the most pronounced colour), which made it possible to rate a blue (hard) tendon as softer if it appeared with yellow/red lesions.

In Phase Two, the adjusted protocol was applied to 20 new participants. Overall agreement in Phase Two corresponded to >80% with blinded raters. Thereafter, Phase Three (n=40) was initiated based on the final protocol, as described above.

Questionnaires

Participants completed questionnaires, including the DASH (Disabilities of the Arm, Shoulder and Hand) for investigating disability of the upper extremities (0-100; 100 being most disabled),[32] The VAS (Visual Analogue Scale) for assessing pain level (0-100; 100 being the most painful),[33] the EQ-5D-3L for measuring health-related quality of life, and the EQ-VAS[34] also for health-related quality of life (0-100; 100 being best imaginable health status). Demographic data included information on age, sex, and BMI.

Statistics

SEL data were found to be normally distributed on a histogram. For continuous data (mean of the three measured areas of the supraspinatus tendon) the ICC (model 2.3, absolute agreement, 2-way random, single measures) was calculated to determine intra-rater and inter-rater reliability.[35] The ICC, with 95% confidence intervals (95% C.I.), was calculated for RAW, and for strain ratios (DELT and GEL). The ICCs were interpreted as <0.40=poor, 0.40-0.59=fair, 0.60-0.74=good and ≥ 0.75=excellent reliability.[36]

Paired student t-tests were completed for statistical comparison of ratings (Rater 1 (first time) vs. Rater 1 (second time), and Rater 1 vs. Rater 2) using a significance level of 0.05. Bland and Altman plots with 95% limits of agreement (LOA) were calculated to evaluate systematic differences,[37] with the 95% LOA calculated as the mean difference ± 1.96 x standard deviation of the difference (SD).[38]

Standard Error of Measurements (SEM) was calculated using the formula: SEM = SDmean difference/√2,[39] and Minimal Detectable Change 95% ($MDC_{individual}$) was calculated using the formula: MDC =SEM x √2 x 1.96[40] and the relative MDC was calculated as a percentage of the average SEL values of Rater 1 and Rater 2.

For ordinal data, linear weighted Cohen's κ (LWk) was used to calculate reliability with 95% CIs for colour ratings and number of lesions. LWk was interpreted as <0=poor agreement, 0.01-0.20=slight agreement, 0.21-0.40=fair agreement, 0.41-0.60=moderate agreement, 0.61-0.80=substantial agreement and >0.81=almost perfect agreement.[41]

Statistical analyses were performed in SPSS, version 25.0 (SPSS Inc., Chicago, IL).

RESULTS

Demographics varied between patients and healthy participants, on most parameters. Due to the sampling method, the patients were expected to be older, with a higher BMI, more pain and disability, and reduced quality of life compared with healthy participants (Table 1).

Table 1
Demographics (mean; SD (frequencies and percentages for dominant arm and EQ-5D)) of patients and healthy participants from the study phase (n = 40).

Variable	**Patients (n = 20)**	**Healthy participants (n = 20)**
Age (years)	47.85 (7.63)	25.70 (6.10)
Females (%)	14 (70)	11 (55)
BMI	30.49 (6.57)	24.94 (2.45)
Dominant arm	12 (60%)	10 (50%)
VAS rest (0-100)	24.60 (20.96)	0 (0.00)
VAS activity (0-100)	53.35 (16.62)	0 (0.00)
VAS sleep (0-100)	49.60 (18.26)	0 (0.00)
VAS maximum (0-100)	78.05 (11.62)	0 (0.00)
DASH (0-100)	34.75 (17.48)	1.71 (4.21)
EQ-VAS (0-100)	53.24 (38.31)	72.90 (36.40)
EQ-5D-3L	**Frequencies**	**Frequencies**
Mobility problems	0 (0%)	0 (0%)
Self-care problems	10 (50%)	1 (5%)
Usual activities problems	1 (5%)	0 (0%)
Pain/discomfort problems	20 (100 %)	1 (5%)
Anxiety/depression problems	5 (20%)	1 (5%)

Abbreviations: SD, Standard Deviation; BMI, Body Mass Index; VAS, Visual Analogue Scale; DASH, Disability of the Arm, Shoulder and Hand Questionnaire; EQ-5D-3L, Quality of Life by dimension; EQ-VAS, Quality of Life

The paired t-test showed statistical differences in inter-rater measurements for RAW and in intra-rater measurements for GEL (Table 2).

Table 2

Reliability of strain elastography in the Supraspinatus tendon using respectively the deltoid muscle (reference) and a gel pad (reference) and raw data from the study phase (n = 40).

Continuous scale	Mean (SD)	Mean (SD)	Mean Difference (SD)	P	LOA	SEM	MDC (%)	ICC (95% C.I.)
Mean across tendon								
Intra-rater	**Rater 1**	**Rater 1**						
RAW	4.40 (0.55)	4.36 (0.55)	0.04 (0.14)	0.09	-0.24; 0.33	0.10	0.28 (6.36)	0.97 (0.93-0.98)
DELT (ratio)	14.33 (3.07)	14.23 (3.16)	0.10 (1.48)	0.66	-2.81; 3.01	1.05	2.91 (20.37)	0.89 (0.80-0.94)
GEL (ratio)	2.87 (1.28)	2.59 (1.01)	0.28 (0.82)	0.04*	-1.33; 1.89	0.58	1.61 (58.82)	0.73 (0.54-0.85)
Inter-rater	**Rater 1**	**Rater 2**						
RAW	4.40 (0.55)	4.53 (0.61)	-0.13 (0.24)	0.00*	-0.60; 0.34	0.17	0.47 (10.53)	0.89 (0.75-0.95)
DELT (ratio)	14.33 (3.07)	14.56 (2.60)	-0.22 (1.88)	0.46	-3.91; 3.46	1.33	3.69 (25.51)	0.78 (0.63-0.88)
GEL (ratio)	2.87 (1.28)	2.73 (1.00)	0.14 (0.89)	0.32	-1.61; 1.90	0.63	1.75 (62.63)	0.70 (0.50-0.83)

Abbreviations: SD, Standard Deviation; LOA, Limits of Agreement; SEM, Standard Error of Mean; MDC, Minimal Detectable Change; ICC (95% C.I.), Intra-class Correlation Coefficient with 95% confidence intervals; DELT, Reference area in the Deltoid Muscle; GEL, Reference area in the Gel Pad; RAW, Raw Elastography Data; *, significant difference ($p \leq 0.05$) between measurements.

None of the Bland and Altman plots showed funnel effects (increasing difference with increasing mean size) (Figure 5).

For intra-rater and inter-rater reliability, there was 'excellent agreement' using RAW and DELT (ICC: intra-rater: 0.97, inter-rater: 0.89), (ICC: intra-rater: 0.89, inter-rater: 0.78) respectively and the reliability was 'good' when using GEL (ICC: intra-rater: 0.73, inter-rater: 0.70) (Table 2). The same pattern of results was also seen for each of the three measured areas of the Supraspinatus tendon (medial, middle and lateral parts) (not shown in tables).

For the intra-rater reliability, the relative MDC was smallest for RAW (6.36%), larger for DELT (20.37%) and largest for GEL (58.82%). For inter-rater reliability, the same pattern was seen for MDC, with a minimum of 10.53%, 25.51% and 62.63%, for RAW, DELT and GEL, respectively.

The reliability of using the colour scale (performed without GEL), kappa (LWk) was very similar, with intra-rater reliability ranging from LWk: 0.76-0.79, representing 'substantial agreement', and inter-rater agreement ranging from LWk: 0.71-0.81 representing 'substantial to almost perfect agreement'.

For the number of yellow/red lesions, LWk was generally highest for intra-rater ranging from 0.40-0.82 representing 'fair to almost perfect agreement', while for inter-rater reliability LWk ranged from 0.24-0.67 representing 'fair to substantial agreement' (Table 3).

Table 3

Reliability of strain elastography in the Supraspinatus tendon using data based on colours, respectively a colour scale and counting the number of lesions from the study phase (n=40)

Ordinal scale	**Intra-rater**		**Inter-rater**	
	Total Agreement (%)	LWκ (95% C.I.)	Total Agreement (%)	LWκ (95% C.I.)
Colourscale				
Medial tendon 1/3	80	0.76 (0.64-0.89)	80	0.77 (0.63-0.91)
Middle tendon 1/3	82.5	0.77 (0.65-0.88)	76	0.71 (0.58-0.83)
Lateral tendon 1/3	85	0.79 (0.65-0.94)	87.5	0.81 (0.65-0.97)
Lesions (no.)				
Medial tendon 1/3	87.5	0.82 (0.69-0.95)	82.5	0.67 (0.53-0.80)
Middle tendon 1/3	90	0.75 (0.51-1.00)	87.5	0.63 (0.30-0.96)
Lateral tendon 1/3	87.5	0.40 (-0.16-0.97)	90	0.24 (-0.14-0.62)

Abbreviations: LWκ (95% C.I.), Linear Weighted κ with 95% confidence intervals

For the medial and middle tendon part of the supraspinatus, LWk was 'substantial to almost perfect' for intra-rater (LWk 0.75-0.82) and 'substantial' for inter-rater (LWk 0.63-0.67) reliability, while for the lateral tendon part, LWk was low, corresponding to only 0.40 and 0.24 for intra-rater and inter-rater reliability, respectively, representing 'fair agreement'.

DISCUSSION

The reliability of the results from SEL was 'excellent' when using the raw data and DELT as reference tissue. When using GEL as reference tissue, the reliability of the results was 'good'. Furthermore, the relative MDC (% of mean) was smallest for RAW and largest for GEL.

When using the colour scale grading, LWk represented 'substantial agreement' to 'substantial to almost perfect agreement'. For the number of yellow/red lesions, LWK was highest for intra-rater reliabiliy, in the medial and middle tendon parts, LWk was 'substantial to almost perfect', while for the lateral tendon part LWk was low, corresponding to only 'fair' agreement'.

Strain Ratios and Raw data

The ROIs with raw SEL data (no reference tissue) resulted in the highest reliability in intra-rater as well as inter-rater reliability. Although there was a significant difference between Rater 1 and Rater 2 for RAW, this difference was less than the MDC and is ascribed to measurement error.

To our knowledge, no study has previously presented raw SEL data, which precludes comparison with other studies. Using raw data has limitations, since there is, in comparison with using ratios, no possibilities to adjust for different transducer pressures. On the other hand, the advantage is that it gives a more quantitative estimate than when using visual inspections, as for example in the colour scale, or with uncertainties due to selection of reference tissue.

In the current study, only one rater captured the images (using almost the same pressure) and two raters traced and measured the areas. If more than one rater were to capture the images, a lower reliability may be expected, due to potentially different transducer pressure.

Different reference tissues in SEL measurements of the musculoskeletal area make it difficult to compare data across studies. The reference tissues previously used for calculating strain ratios in the achilles and supraspinatus tendon areas have included bone,[15] fat,[13 25 31 42 43] and GEL/acoustic coupler.[13 16 44] Using bone as a reference value has limitations, since ultrasound cannot penetrate bone material, meaning that data coming from this region are artefacts. In addition, strain ratios may be based on different equations in different studies, placing the ROIs of the tendon in the denominator,[13 45] or, as in the currrent study, in the numerator.[25 31] Due to thin subcutaneous fat tissue area in some participants, the reference tissue was replaced, after Phase One, with DELT and GEL (artificial fat tissue).
When using DELT as the reference tissue, reliability was found to be excellent (both intra-rater and inter-rater). Muscle tissue increases stiffness significantly after exercise and muscle contractions,[16 24] and therefore, limitations are recognised when using this tissue as reference when for example, investigating tendon tissue response to muscle contractions. However, no previous study of SEL on the supraspinatus tendon has used muscle tissue as a reference for the calculated strain ratios.

The current study found a high MDC for GEL (large measurement error) but a good reliability (ICC 0.70-0.73), in line with previous studies. One study, using the Kager's fat pad (fat deposit within the Kager triangle anterior to the achilles tendon) as reference tissue in the

achilles tendon, also found good to excellent intra-rater and inter-tester reliability (ICC: 0.51-0.78) in the longitudinal plane; however, with much lower reliability (ICC: 0.41-0.45) in the axial plane.[25] Unfortunately, MDC was not reported. The ICC and MDC in the current study are reported as single measures, as it is of relevance to clinicians. None of the previous studies[16 44] have described whether ICC is reported as a single measure or (group) average measure, however, it is anticipated that it is a group average measure. In the current study, the ICC will increase about 10% and the MDC will decrease up to 80% when reported as average measures. This could explain some of the differences in results between the current and previous studies.

In the current study, the data using GEL are also in line with a study using an acoustic coupler as reference tissue in the supraspinatus tendon,[16] where excellent intra-rater reliability was shown, as was also confirmed for the achilles tendon.[44] An acoustic coupler is similar to GEL and may be acceptable, but ideally, the reference area should be the same depth as the ROI (in this case, the supraspinatus tendon). As GEL is not located close to the tendon, the ROIs using GEL will not be subjected to the same amount of tissue pressure as the tendon, which may affect reliability and validity.

The current study found the lowest reliability (however, it was still graded 'good') when using GEL which may be caused by difficulties locating the footprint of the tendon, due to lower image quality, because of increased depth (through 10 mm of GEL).

A statistically significant intra-rater difference was also seen, but as this difference was below the MDC, it can be ascribed to measurement error.

Colour Ratings

The current study found a high reliability ('substantial' to 'almost perfect') when grading the ROIs of SEL images according to a 4-level colour scale, where blue tissue indicated hard tissue and red tissue indicated soft tissue. This is in line with a previous study on the achilles tendon using colour scales of 5 levels (1=blue (hardest tissue), 2=light blue, 3=green, 4=yellow, 5=red (softest tissue)), where good to excellent inter-rater (κ = 0.897) and intra-rater (κ = 1) reliability was found.[30]

Alternative types of colour scales have been used in the musculoskeletal area, primarily for the achilles tendon, with categorisation of the achilles tendon into a two-level category scale (green/blue vs. red),[25] and a three-level category scale (blue/yellow/red).[22 29] The current study has used the same colours for defining hard tissue (blue), and for soft tissue (red), but with a four-level category scale, and with more precise criteria for the different levels which were found necessary due to the mixture of colours in the supraspinatus tendon observed in Phase One.

In the current study, LWk in the medial and the middle parts of the tendon was 'substantial to almost perfect' when counting the number of yellow/red lesions in both intra-rater and inter-rater reliability. Even though the total agreement was also high in the lateral third of the tendon, LWk was relatively low compared with the medial and the middle parts of the tendon. The reason may be due to the low presence of lesions in the lateral part, corresponding to only 10 % of the participants presenting with lesions in the lateral 1/3 of the tendon, which can lead to the 'Kappa Paradox'.[46] The Kappa Paradox means (in a 2x2 table) that an imbalance between presence and absence in overall agreement, and between

disagreement (bias) can drastically lower the kappa, why prevalence and bias-adjusted kappa (PABAK) in dichotomous scales is recommended.[47] A PABAK on the current number of lesions dichotomised into no lesion /lesion(s) will increase the kappa by up to 70%. Another explanation is that since the pressure is placed vertically on the medial part of the tendon, the lateral part will have experienced a smaller degree of stress.

The high reliability of the number of focal lesions (LwK intra-rater reliability: 87.5-90, LwK inter-rater reliability: 82.5-90) is in line with a previous study on the supraspinatus tendon where an almost prefect reliability (k = 0.83) was found.[14]

The current methods of using colour grading and lesion counts for assessing tendon stiffness in the supraspinatus tendon showed high reliability. These methods are feasible for use in clinical practice as they can be performed quickly, but when more than one clinician is performing the SEL, the method cannot adjust for potentially different tranducer pressures.

Limitations

Manual compression may affect reliability, especially when using raw data, colour grading and counting yellow/red lesions. To partly counteract this, a quality bar was used that provided instant feedback on the uniformity of the transducer pressure. Reference tissues (DELT and GEL) with calculation of strain ratios were also used, thereby making a comparison possible between the methods.

Furthermore, as mentioned, SEL is highly operator-dependent, why the same (and only one) trained operator captured the present images, thereby further decreasing the risk of bias, but this limits the external validity of the results.

In addition, age was not blinded and, as reported earlier, tendons get softer with age.[48] Therefore, the current blinding of examiner results and health status may not have been sufficient.

Strengths

The strength of this study was the design, incorporating a stepwise and standardised procedure for reliability which minimised bias and increased reliability.[17] Furthermore, both patients and healthy participants were enrolled.

To enhance standardisation, all SEL ROIs were measured at a fixed point, just laterally from the anterior-lateral corner of the acromion in the longitudinal plane of the supraspinatus tendon.[20] In addition, the reliability was estimated by using three different quantitative methods (RAW, DELT and GEL), as well as two different qualitative methods (colour and number of yellow/red focal lesions).

CONCLUSION

Intra-rater and inter-rater reliability were excellent using raw values and ratios with the deltoid muscle as reference tissue, and good when using a gel pad as reference tissue. The reliability of colour scale ratings was substantial to almost perfect and the number of lesions was fair to almost perfect.

Although high reliability was found, validity and responsiveness of these elastographic methods needs further investigation.

Contributorship statement

KB, JH, PK, KGI and BJ-K conceived and designed the study protocol. KB and BJ-K procured the project funding. KB, JH, PK, and BJ-K developed and standardised the ultrasound procedure. JH, KB and KGI recruited participants.

KB was the project coordinator and captured the strain elastography images. KGI performed the physical tests. KB and JH rated the images. KB and BJ-K planned and coordinated the statistical analyses. KB performed the statistical analyses. KB drafted the manuscript, and JH, PK, KGI and BJ-K contributed to the manuscript. All authors read and approved the final manuscript. KB is the guarantor.

Competing interests

The authors declare no conflicts of interest.

Funding

This work was supported by the Region of South Denmark (14/38969, 15/24799), The Danish Rheumatism Association (R150-A4296) and the Danish Council of Radiographers (no grant number).

Data sharing statement

No additional data are available.

REFERENCES

1. Luime JJ, Koes BW, Hendriksen IJ, et al. Prevalence and incidence of shoulder pain in the general population; a systematic review. *Scand J Rheumatol* 2004;33(2):73-81.

2. Singh SG, Sumit.; Mohammad, F.; Kumar, S.; Kumar, D.; Kumar, S. Prevalence of shoulder disorders in tertiary care centre. *J Res Med Sci* 2015;3(4):917-20.

3. Nyman PP, K.; Panula, H.; Mälkiä, E. Patients' experience of shoulder problems prior to and following intervention. *Physiother Theory Pract* 2012;28(3):221-31.

4. Virta L, Joranger P, Brox JI, et al. Costs of shoulder pain and resource use in primary health care: a cost-of-illness study in Sweden. *BMC Musculoskelet Disord* 2012;13:17.

5. Teunis T, Lubberts B, Reilly BT, et al. A systematic review and pooled analysis of the prevalence of rotator cuff disease with increasing age. *J Shoulder Elbow Surg* 2014;23(12):1913-21.

6. Klauser AS, Faschingbauer R, Jaschke WR. Is sonoelastography of value in assessing tendons? *Semin Musculoskelet Radiol* 2010;14(3):323-33.

7. Prado-Costa R, Rebelo J, Monteiro-Barroso J, et al. Ultrasound elastography: compression elastography and shear-wave elastography in the assessment of tendon injury. *Insights Imaging* 2018

8. Dietrich CF, Barr RG, Farrokh A, et al. Strain Elastography - How To Do It? *Ultrasound Int Open* 2017;3(4):E137-e49.

9. Garra BS. Elastography: history, principles, and technique comparison. *Abdom Imaging* 2015;40(4):680-97.

10. Chung SW, Kim JY, Yoon JP, et al. Arthroscopic repair of partial-thickness and small full-thickness rotator cuff tears: tendon quality as a prognostic factor for repair integrity. *Am J Sports Med* 2015;43(3):588-96.

11. Bamber J, Cosgrove D, Dietrich CF, et al. EFSUMB guidelines and recommendations on the clinical use of ultrasound elastography. Part 1: Basic principles and technology. *Ultraschall Med* 2013;34(2):169-84.

12. Fusini F, Langella F, Busilacchi A, et al. Real-time sonoelastography: principles and clinical applications in tendon disorders. A systematic review. *Muscles Ligaments Tendons J* 2017;7(3):467-77.

13. Lee SU, Joo SY, Kim SK, et al. Real-time sonoelastography in the diagnosis of rotator cuff tendinopathy. *J Shoulder Elbow Surg* 2016;25(5):723-9.

14. Seo JB, Yoo JS, Ryu JW. Sonoelastography findings of supraspinatus tendon in rotator cuff tendinopathy without tear: comparison with magnetic resonance images and conventional ultrasonography. *J Ultrasound* 2015;18(2):143-9.

15. Tudisco C, Bisicchia S, Stefanini M, et al. Tendon quality in small unilateral supraspinatus tendon tears. Real-time sonoelastography correlates with clinical findings. *Knee Surg Sports Traumatol Arthrosc* 2013

16. Muraki T, Ishikawa H, Morise S, et al. Ultrasound elastography-based assessment of the elasticity of the supraspinatus muscle and tendon during muscle contraction. *J Shoulder Elbow Surg* 2015;24(1):120-6.

17. Patijn J. Reproducibility and Validity Studies of Diagnostic Procedures in Manual/Musculoskeletal Medicine 2004.

18. Kottner J, Audige L, Brorson S, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol* 2011;64(1):96-106.

19. Hougs Kjaer BPM, Ellegaard KPP, Wieland IP, et al. Intra-rater and inter-rater reliability of the standardized ultrasound protocol for assessing subacromial structures. *Physiother Theory Pract* 2017;33(5):398-409.

20. Ingwersen KG, Hjarbæk J, Eshøj H, et al. Ultrasound assessment for grading structural tendon changes in supraspinatus tendinopathy: an inter-rater reliability study. *B M J Open* 2016;6(5)

21. Sein ML, Walton J, Linklater J, et al. Reliability of MRI assessment of supraspinatus tendinopathy. *Br J Sports Med* 2007;41(8):e9.

22. De Zordo T, Fink C, Feuchtner GM, et al. Real-time sonoelastography findings in healthy Achilles tendons. *AJR Am J Roentgenol* 2009;193(2):W134-8.

23. Cosgrove D, Piscaglia F, Bamber J, et al. EFSUMB guidelines and recommendations on the clinical use of ultrasound elastography. Part 2: Clinical applications. *Ultraschall Med* 2013;34(3):238-53.

24. Niitsu M, Michizaki A, Endo A, et al. Muscle hardness measurement by using ultrasound elastography: a feasibility study. *Acta Radiol* 2011;52(1):99-105.

25. Drakonaki EE, Allen GM, Wilson DJ. Real-time ultrasound elastography of the normal Achilles tendon: reproducibility and pattern description. *Clin Radiol* 2009;64(12):1196-202.

26. Lumsdaine W, Smith A, Walker RG, et al. Morphology of the humeral insertion of the supraspinatus and infraspinatus tendons: Application to rotator cuff repair. *Clin Anat* 2015;28(6):767-73.

27. Curtis AS, Burbank KM, Tierney JJ, et al. The insertional footprint of the rotator cuff: an anatomic study. *Arthroscopy* 2006;22(6):609.e1.

28. De Zordo T, Lill SR, Fink C, et al. Real-time sonoelastography of lateral epicondylitis: comparison of findings between patients and healthy volunteers. *AJR Am J Roentgenol* 2009;193(1):180-5.

29. Klauser AS, Miyamoto H, Tamegger M, et al. Achilles tendon assessed with sonoelastography: histologic agreement. *Radiology* 2013;267(3):837-42.

30. Sconfienza LM, Silvestri E, Cimmino MA. Sonoelastography in the evaluation of painful Achilles tendon in amateur athletes. *Clin Exp Rheumatol* 2010;28(3):373-8.

31. Turan A, Tufan A, Mercan R, et al. Real-time sonoelastography of Achilles tendon in patients with ankylosing spondylitis. *Skeletal Radiol* 2013;42(8):1113-8.

32. Kitis A, Celik E, Aslan UB, et al. DASH questionnaire for the analysis of musculoskeletal symptoms in industry workers: a validity and reliability study. *Appl Ergon* 2009;40(2):251-5.

33. Boonstra AM, Schiphorst Preuper HR, Reneman MF, et al. Reliability and validity of the visual analogue scale for disability in patients with chronic musculoskeletal pain. *Int J Rehabil Res* 2008;31(2):165-9.

34. Norman R, Cronin P, Viney R, et al. International comparisons in valuing EQ-5D health states: a review and analysis. *Value Health* 2009;12(8):1194-200.

35. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;86(2):420-8.

36. Cicchetti D. Guidelines, Criteria, and Rules of Thumb for Evaluating Normed and Standardized Assessment Instrument in Psychology. *Psychol Assess* 1994;6(4):284-90.

37. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1(8476):307-10.

38. Ottenbacher KJ, Stull GA. The analysis and interpretation of method comparison studies in rehabilitation research. *Am J Phys Med Rehabil* 1993;72(5):266-71.

39. de Vet HCW, Terwee CB, Knol DL, et al. When to use agreement versus reliability measures. *J Clin Epidemiol* 2006;59(10):1033-9.

40. de Vet HCW, Terwee CB, Knol DL, et al. Measurements in Medicine - a pratical guide. New York, USA: Cambridge University Press 2011.

41. Landis RJK, G.G. The Measurement of Observer Agreement for Categorical Data. *Biometrics* 1977;33(1):159–74.

42. Busilacchi A, Olivieri M, Ulisse S, et al. Real-time sonoelastography as novel follow-up method in Achilles tendon surgery. *Knee Surg Sports Traumatol Arthrosc* 2014

43. Ooi CC, Schneider ME, Malliaras P, et al. Diagnostic performance of axial-strain sonoelastography in confirming clinically diagnosed Achilles tendinopathy: comparison with B-mode ultrasound and color Doppler imaging. *Ultrasound Med Biol* 2015;41(1):15-25.

44. Yamamoto Y, Yamaguchi S, Sasho T, et al. Quantitative Ultrasound Elastography With an Acoustic Coupler for Achilles Tendon Elasticity: Measurement Repeatability and Normative Values. *J Ultrasound Med* 2016;35(1):159-66.

45. Akkaya S, Akkaya N, Agladioglu K, et al. Real-time elastography of patellar tendon in patients with auto-graft bone-tendon-bone anterior cruciate ligament reconstruction. *Archives of Orthopaedic and Trauma Surgery* 2016;136(6):837-42.

46. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol* 1990;43(6):543-9.

47. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *J Clin Epidemiol* 1993;46(5):423-9.

48. Hsiao MY, Chen YC, Lin CY, et al. Reduced patellar tendon elasticity with aging: In vivo assessment by shear wave elastography. Ultrasound Med. Biol. 2015;41(11):2899-905.

Figure legends

Figure 1

With the arm located behind the back and the elbow flexed at 90° and the palm facing posteriorly, the probe was placed just laterally from the anterior-lateral corner of the acromion in the longitudinal plane of the supraspinatus tendon.

Figure 2

Left side: For measuring elastography characteristics of the supraspinatus tendon, the tendon was split into three parts (3 x 7.7mm), illustrated by the areas of blue, red and green colours. This division was based on a line from the lateral tendon insertion (greater tubercle) to the medial tendon insertion, corresponding to 6.5 mm, and from there a line of 23 mm to the medial tendon with the end point (medial part) being perpendicular to the superior surface of the tendon.

The yellow circle in the soft part of the deltoid muscle is used as the reference tissue.

Right side: The three measurement areas (ROIs) and one reference area with elastography characteristics (raw data) during the time of measurement.

Figure 3

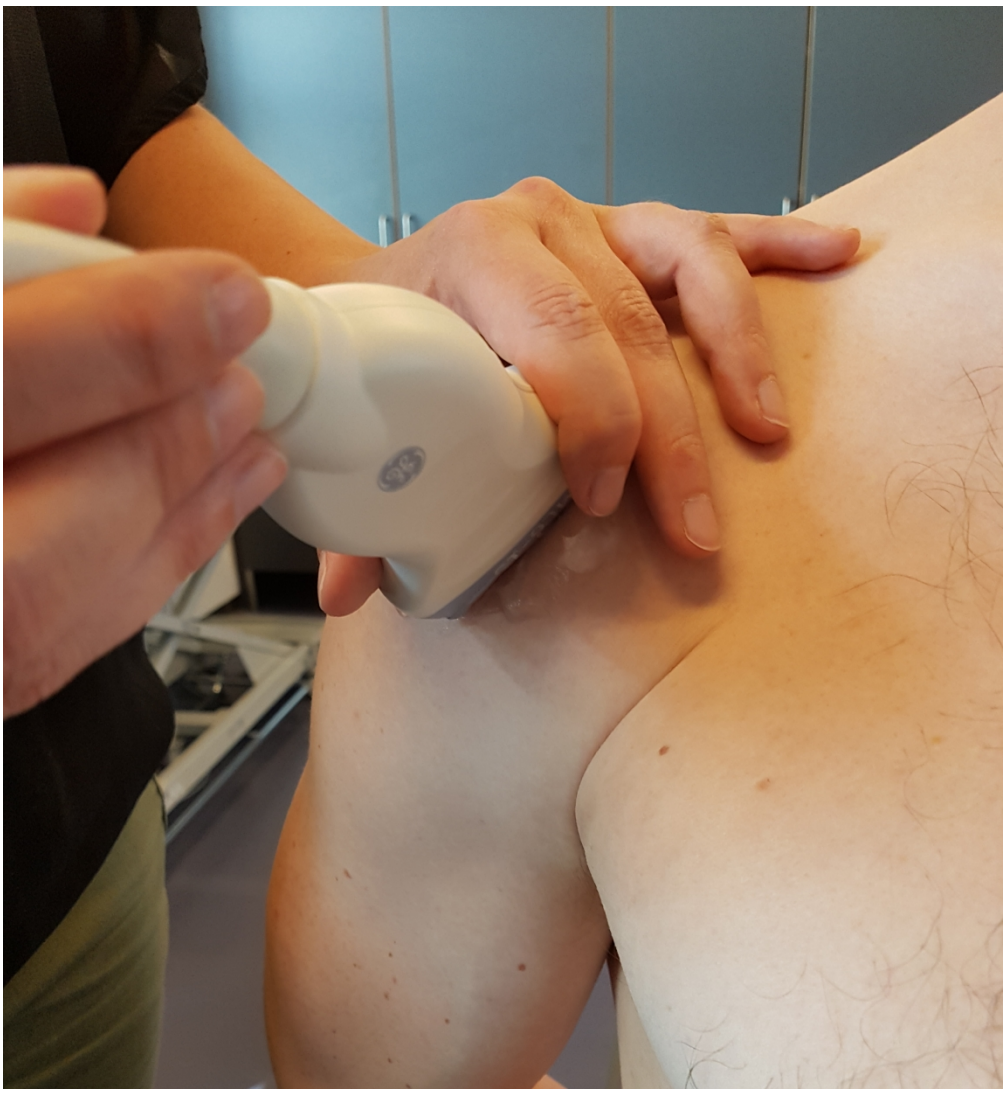
Left side: Gel pad (mounted on the transducer with a condom used as the other reference area, yellow circle). Right side: The corresponding elastography image/measurements.

Figure 4

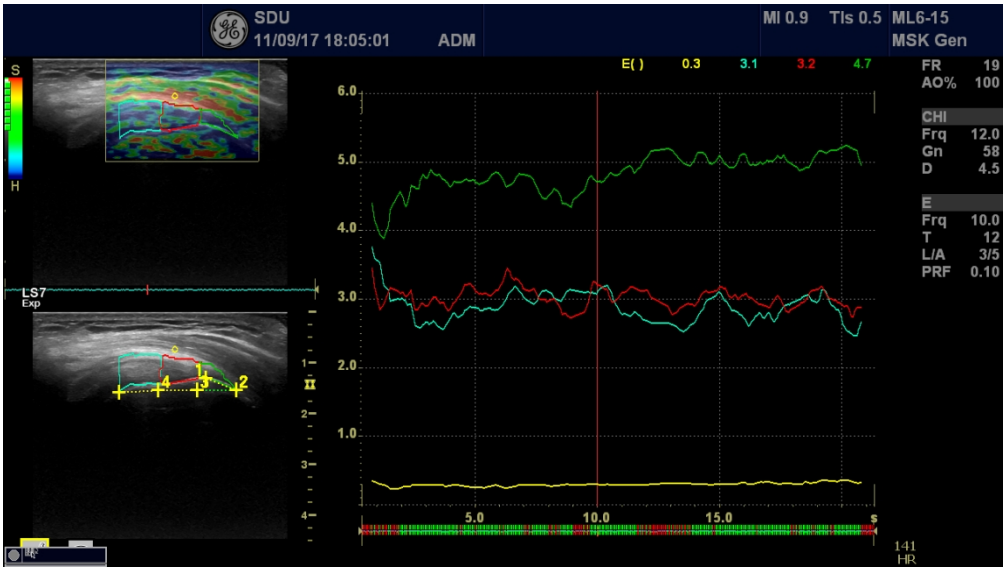
A 6.5 mm chord was drawn (in a lateral direction) from the medial part of the insertion on the humeral head to the lateral part of the tuberculum major. This fixed point (end of the line at the lateral part of the insertion) was used to draw a 23 mm (7.7 x 3) horizontal line in a medial direction ensuring agreement of measuring area.

Figure 5

Bland-Altman plots with 95% Limits of Agreement for measurements, using respectively raw data, the deltoid muscle and a gel pad as reference areas. Values are based on the mean of the three measured parts across the supraspinatus tendon.

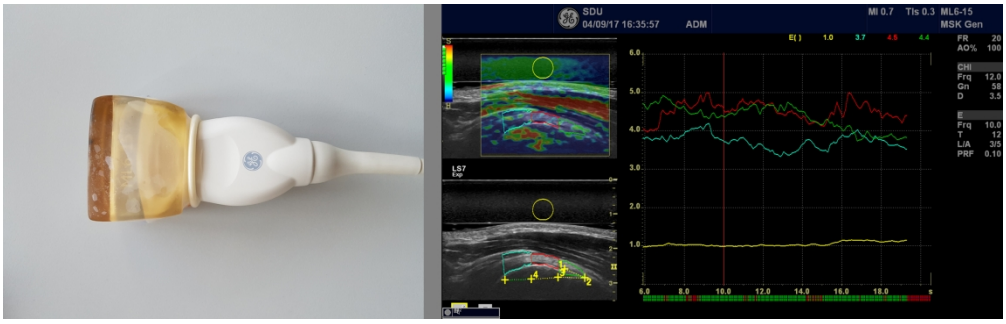


With the arm located behind the back and the elbow flexed 90° with the palm facing towards the posterior direction, the probe is placed just laterally from the anterior-lateral corner of the acromion in the longitudinal plane of the supraspinatus tendon.

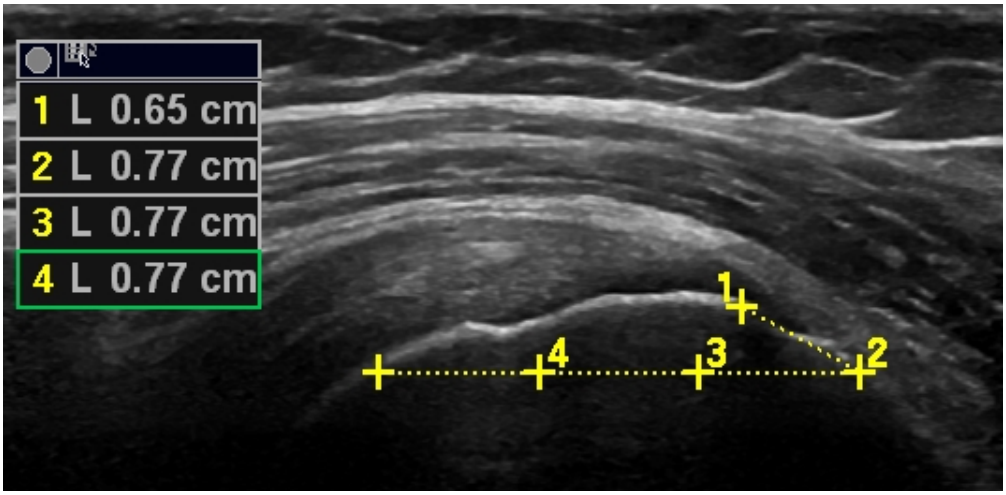


Left side: For measuring elastography characteristics of the supraspinatus tendon, the tendon was split into 3 parts (3x7.7mm), illustrated by the areas of blue, red and green colors. This division was based on a line from the lateral tendon insertion (tuberculum majus) to the medial tendon insertion part, corresponding to 6.5 mm, and from there a line of 23 mm to the medial tendon part with the end point (medial part) being perpendicular to the superior surface of the tendon.
The yellow circle in the soft part of the deltoid muscle is used as a reference.
Right side: The three measurements areas (ROI's) and one reference area with elastography characteristics (raw data) during the time of measurement.

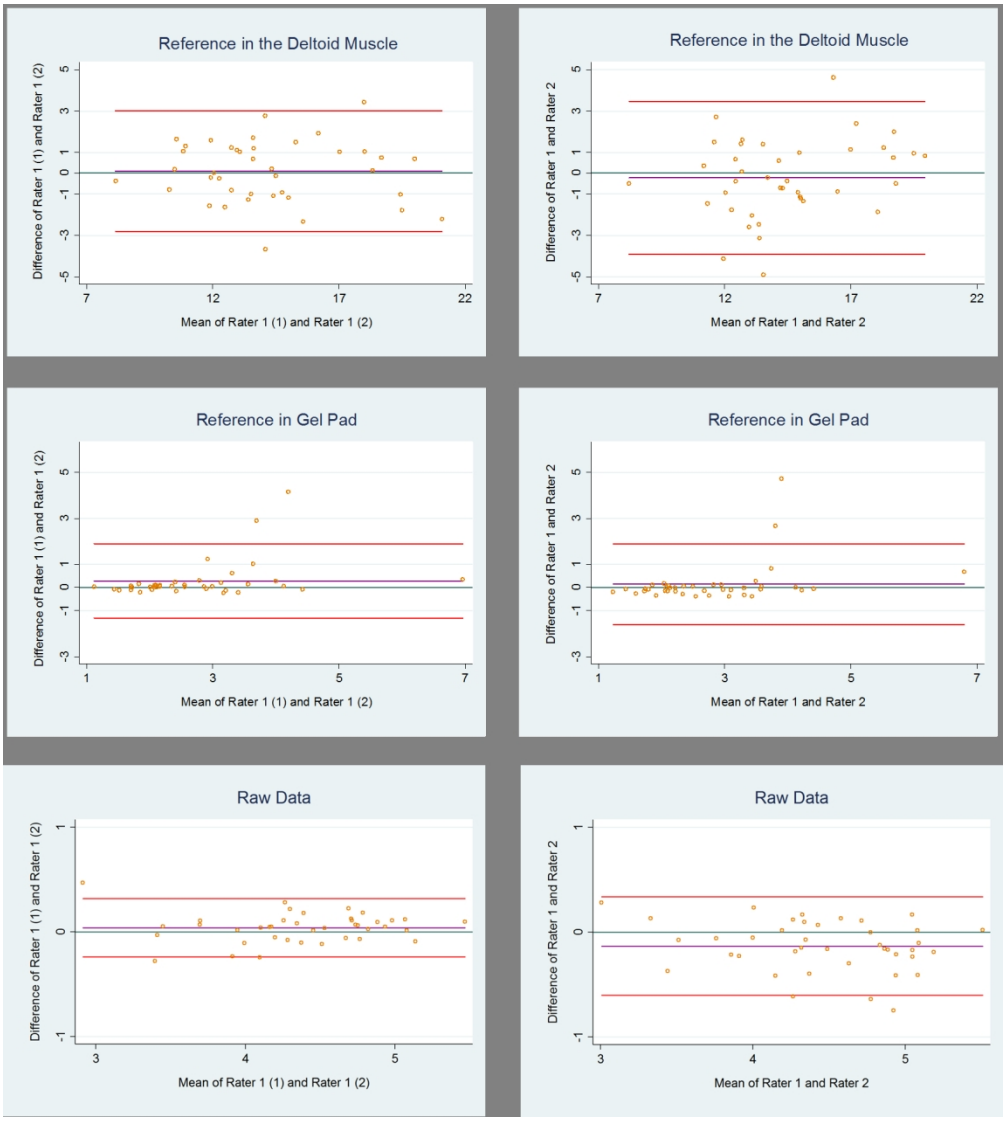
451x254mm (72 x 72 DPI)



Left side: Gel pad (mounted on the transducer with a condom used as the other reference area, yellow circle). Right side: The corresponding elastography image/measurements.



A 6.5 mm chord was drawn (in the lateral direction) from the medial part of the insertion on the humeral head to the lateral part of the tuberculum major. This fix point (end of the line at the lateral part of the insertion) was used to draw a 23 mm (7.7 x 3) horizontal line in the medial direction ensuring agreement of measuring area.



Bland-Altman plots with 95% Limits of Agreement for measurements, using respectively raw data, the deltoid muscle and a gel pad as reference areas. Values are based on the mean of the three measured parts across the supraspinatus tendon.

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract **(p. 1, l. 1)**
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found **(p. 2, l. 1)**
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported **(p. 4, l. 1)**
Objectives	3	State specific objectives, including any prespecified hypotheses **(p. 6, l. 11)**
Methods		
Study design	4	Present key elements of study design early in the paper **(p. 7, l. 2)**
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection **(p. 7, l. 14)**
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up **(p. 8, l. 1)**
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable **(p. 10, l. 19)**
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group **(p. 10, l. 19)**
Bias	9	Describe any efforts to address potential sources of bias **(p. 6, l. 1 & (p. 11, l. 16)**
Study size	10	Explain how the study size was arrived at **(p. 4, l. 5)**
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why **(p. 13, l. 16)**
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding
		(*b*) Describe any methods used to examine subgroups and interactions
		(*c*) Explain how missing data were addressed
		(*d*) If applicable, explain how loss to follow-up was addressed
		(*e*) Describe any sensitivity analyses
Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed
		(b) Give reasons for non-participation at each stage
		(c) Consider use of a flow diagram
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders **(p. 15, l. 5)**
		(b) Indicate number of participants with missing data for each variable of interest
		(c) Summarise follow-up time (eg, average and total amount)
Outcome data	15*	Report numbers of outcome events or summary measures over time
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included
		(*b*) Report category boundaries when continuous variables were categorized

		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses
Discussion		
Key results	18	Summarise key results with reference to study objectives **(p. 20, l. 1)**
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias **(p. 24, l. 16)**
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence **(p. 20, l. 18)**
Generalisability	21	Discuss the generalisability (external validity) of the study results
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based **(p. 27, l. 1)**

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.

BMJ Open

ULTRASONIC STRAIN ELASTOGRAPHY FOR DETECTING ABNORMALITIES IN THE SUPRASPINATUS TENDON: AN INTRA- AND INTER-RATER RELIABILITY STUDY

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-027725.R2
Article Type:	Research
Date Submitted by the Author:	09-Apr-2019
Complete List of Authors:	Brage, K; University of Southern Denmark, Sport Science and Clinical Biomechanics Hjarbaek, John; Odense Universitetshospital, Department of Radiology, Musculoskeletal section Kjaer, Per ; University of Southern Denmark, Institute of Sports Science and Clinical Biomechanics Ingwersen, Kim G.; Sports Science and Clinical Biomechanics Juul-Kristensen, Birgit; Research Unit for Musculoskeletal Function and Physiotherapy, Institute of Sports Science and Clinical Biomechanics
Primary Subject Heading:	Radiology and imaging
Secondary Subject Heading:	Diagnostics
Keywords:	Strain elastography, supraspinatus tendon, rotator cuff, tendon quality, reliability, tendinopathy

ULTRASONIC STRAIN ELASTOGRAPHY FOR DETECTING ABNORMALITIES IN THE SUPRASPINATUS TENDON: AN INTRA- AND INTER-RATER RELIABILITY STUDY

Brage K[1,2], Hjarbaek J[3], Kjaer P[1,2], Ingwersen KG[4], Juul-Kristensen B[1]

[1] Department of Sports Science and Clinical Biomechanics, University of Southern Denmark, Campusvej 55, 5230 Odense M, DK
[2] UCL University College, Niels Bohrs Allé 1, 5230, Odense M, DK
[3] Department of Radiology, Odense University Hospital, J. B. Winsløws Vej 4, 5000 Odense C, DK
[4]Research Unit in Physiotherapy and Occupational Therapy, Hospital Lillebaelt, Kabbeltoft 25, 7100 Vejle, DK

Author information
Corresponding Author
Karen Brage, PhD candidate, Telephone: +45 51787381, fax: 6318 3226, e-mail: kabr2@ucl.dk
Co-Authors
John Hjarbaek, MD, Telephone: +45 66113333, e-mail: john.hjarbaek@rsyd.dk
Per Kjaer, Professor, PhD., Telephone: +45 65504553, e-mail: pkjaer@health.sdu.dk
Kim Gordon Ingwersen, PhD., Telephone: +45 65509044, e-mail: kingwersen@health.sdu.dk
Birgit Juul-Kristensen, PhD.: Telephone: +45 65503412, e-mail: bjuul-kristensen@health.sdu.dk

Wordcount
4404

Trial registration
The study protocol was approved by the Ethics Committee for the Region of South Denmark (S-20160115) and reported to the Danish Data Protection Agency (2014-41-3266).

ABSTRACT

Objectives

The reliability of ultrasonic strain elastography (SEL) used to detect abnormalities in the supraspinatus tendon is unclear. Thus, the aim of this study was to investigate the reliability of SEL in the supraspinatus tendon.

Design

An intra-rater and inter-rater reliability study.

Setting

A single-centre study conducted at the University of Southern Denmark.

Participants

Twenty participants with shoulder pain and MRI-verified supraspinatus tendinosis and 20 asymptomatic participants (no MRI).

Primary and secondary outcome measures

Raw values (RAW), and ratios (deltoid muscle (DELT); gel pad (GEL) as reference tissues) were calculated and mean values of measurements from three regions of the supraspinatus tendon were reported. Colour scale ratings and number of yellow/red lesions from the three areas were also included.

Results

Intra-rater reliability showed Intra-Class Correlation Coefficients (ICCs) for RAW, DELT and GEL: 0.97 (Minimal Detectable Change (MDC): 0.28 (6.36% of the mean); 0.89 (MDC: 2.91 (20.37%); and 0.73 (MDC: 1.61 (58.82%)), respectively. The ICCs for inter-rater reliability were 0.89 (MDC: 0.47 (10.53%); 0.78 (MDC: 3.69 (25.51%); and 0.70 (MDC: 1.75 (62.63%)), respectively.

For colour scale ratings, intra-rater reliability (Linear Weighted kappa, LWk) ranged from 0.76 to 0.79, with the inter-rater reliability from 0.71 to 0.81. For the number of lesions, intra-rater reliability ranged from 0.40 to 0.82, and inter-rater reliability from 0.24 to 0.67.

Conclusions

Intra-rater and inter-rater reliability were excellent for raw values and for ratios with deltoid muscle as the reference tissue, and good for ratios with gel pad as the reference tissue. The reliability of colour scale ratings was substantial to almost perfect, and for the number of lesions fair to almost perfect.

Although high reliability was found, validity and responsiveness of these elastographic methods needs further investigation.

Strengths and limitations of this study

- Standardised procedure for capturing strain elastographic images was developed and applied
- Raw values of strain elastographic images were presented
- Ratios, based on different reference tissues or areas (deltoid muscle/gel pad), of strain elastographic images were calculated and presented
- Specific procedures for colour grading strain elastographic images were presented
- SEL is highly operator-dependent which may limit the external validity

INTRODUCTION

Shoulder pain is a common symptom, with a lifetime prevalence in the general population of 6.7-66.7%[1] and subacromial pain syndrome is the second most common cause of pain in the shoulder.[2] Shoulder pain has consequences such as physical limitations, mental problems[3] and absence from work.[4]

Shoulder disorders are evaluated by history-taking and physical examination, potentially supplemented with X-ray, conventional ultrasound and/or Magnetic Resonance Imaging (MRI). However, supraspinatus tendon abnormalities are also found in asymptomatic people when using general modalities such as MRI and ultrasound.[5] Furthermore, it seems difficult to distinguish pathological changes from healthy tissues by using conventional ultrasound because pathological regions often exhibit the same echogenicity as non-pathological regions.[6]

Strain elastography (SEL) is a relatively new and not yet well-established method, which may assist in diagnosis, prediction and monitoring of progress in tendon healing.[7]

SEL defines the physical properties of soft tissues through characterisation of the differences in stiffness between 'the regions of interest' (ROI) and the surrounding tissues.[8 9] Conventional ultrasound and MRI were developed for visualisation of macroscopic changes and not specifically for the mechanical tendon properties, which is why SEL may add further knowledge to conventional shoulder imaging. Since tendon quality is a prognostic factor for rotator cuff repair, information about tendon stiffness could be beneficial for the surgeon.[10]

Tissue deformation using SEL is obtained by uniform mechanically induced compressions (strain) of the structures under the ultrasound transducer, during the ultrasound scan. Through manual

compression, the soft tissue deforms differently depending on its inherent stiffness. The degree of deformation can be interpreted as an estimate of the tissue stiffness.[11]

SEL has recently been used in the musculoskeletal area, where the achilles tendon has been the primary focus.[12] A few studies have found significant associations between pathology identified by SEL and MRI in patients with supraspinatus tendinopathy,[13 14] besides significant correlations between results from SEL and clinical tests and questionnaires in patients with small supraspinatus tendon tears.[15] One study also found SEL to be able to detect increased stiffness in the supraspinatus tendon elasticity and muscle elasticity with increased muscle contraction in healthy participants.[16] Although only a few studies have been performed, and with different aims, comparator instruments, procedures, reference tissues and data types, the concurrent validity of SEL in the supraspinatus tendon seems promising.[13-16]

However, the reliability of SEL must first be proven to be acceptable. In this respect, SEL constitutes several challenges since SEL is a technique with relatively high operator dependency in terms of the manually applied pressure and subsequent identification and selection of the pathological ROI.

One of only two reliability studies found the inter-rater reliability of SEL in the supraspinatus tendon to be almost perfect (κ = 0.83) with respect to the number of focal lesions in 118 patients with MRI-verified supraspinatus tendinopathy.[14] However, this study did not include a healthy control group which is recommended for reliability and validity studies, to achieve realistic tissue variation.[17] Furthermore, the study used only colour quantification.[14] The other study found a high intra-rater reliability (ICC $_{1,3}$ = 0.92-0.99) with respect to ROI, when using an acoustic coupler as

the reference tissue in a small sample of 23 healthy participants. The limitations of this study were: not including a group with pathology, not defining ROI and only using one type of reference tissue (acoustic coupler).[16] Since standardised and consensus procedures for conducting elastography in the supraspinatus tendon have not yet been established, there is a need to investigate which reference tissue has the highest reliability. To our knowledge, there has not been a study investigating the reliability of SEL in the supraspinatus tendon that has included both patients with a pathological (non-ruptured) supraspinatus tendon and healthy participants with non-painful shoulders. Furthermore, choice of reference tissue and quantification methods has major impact on results, but no reliability studies have compared these different approaches.

The aim of this study was to test the intra-rater and inter-rater reliability of SEL within the supraspinatus tendon in patients and healthy participants, using different reference tissues (deltoid muscle (DELT), gel pad (GEL)), and different quantification methods (raw data, strain ratios, colour scale rating and counting number of yellow/red lesions).

MATERIALS AND METHODS

This study was an intra-rater and inter-rater reliability study of SEL, used on the supraspinatus tendon, reported according to the Guidelines for Reporting Reliability and Agreement Studies (GRRAS).[18]

A three-phased intra-rater and inter-rater reliability protocol for diagnostic reliability studies was used.[17] The protocol included a training phase (Phase One), an overall agreement phase (Phase Two), and an actual study phase (Phase Three), for securing low clinician dependency and subjectivity, sufficient experience and standardisation, and minimisation of systematic bias.[17] In order to progress to the study phase, the criterion of at least 80% inter-rater agreement in Phase Two was used. Such a protocol has previously been used in reliability studies, using ultrasound methods.[19 20]

Study procedures

The study's participants were recruited from August 2016 to December 2017. Symptomatic participants, with an MRI-verified supraspinatus tendinosis (*patients)*, were recruited from the Radiology Department at Odense University Hospital within 14 days of their MRI examination. Participants with no shoulder symptoms (*healthy participants)* were recruited primarily from the University of Southern Denmark and the UCL University College, both situated in Odense. Except for MRI, all the procedures were performed at the University of Southern Denmark. After inclusion, participants underwent clinical tests (performed by KGI) and filled out questionnaires regarding functional limitations, pain and quality of life. Testing procedures took place once and lasted approximately one hour.

All SEL images were captured by the same clinician (KB), while two raters performed the SEL measurements on the captured images. Rater 1 (KB), radiographer, was thoroughly trained in using SEL on the shoulder by Rater 2 (JH), radiologist, who had more than 20 years' experience in clinical musculoskeletal ultrasound. In Phase Three, the raters were blinded to each other's results (data were stored separately), the participants' health status (Rater 1 entered the room after clinical tests and questionnaires were performed, and Rater 2 had no contact with the participants) and their MRI results (ID-numbers were changed after MRI examinations).

SEL images were stored for at least 14 days after image capturing until the first image assessment, to ensure elimination of any memory of pain response during SEL by Rater 1. Further, all SEL images were stored for an additional 14 days before Rater 1's second assessment, to ensure elimination of recalls of SEL results from the first assessment.

The study protocol was approved by the Ethics Committee for the Region of South Denmark (S-20160115) and reported to the Danish Data Protection Agency (2014-41-3266). All participants had oral and written information about the study and signed an informed consent form before participation in this study.

Patient and Public Involvement

Patients and the public were not involved in the design or planning of the study.

Participants

Inclusion and exclusion criteria

An inclusion criterion for both patients and healthy participants was age 18-65 years. In addition, patients had to have at least one shoulder clinically diagnosed with tendinosis (positive signs from ≥ 3 clinical tests (Hawkins-Kennedy test; Neers test; Empty Can test; Full Can test; Resisted external rotation test)) and based on MRI (≥ Grade 1; corresponding to focal increase in tendon signal on protondensity-weighted and fat-suppressed T2 sequences not equal to that of fluid.[21] If both shoulders were MRI-scanned, the most severely affected side was included.

Inclusion criteria for the healthy participants were no previously experienced shoulder problems, and negative signs from all five clinical tests described above. For these participants, the choice of shoulder was matched with that of the patients.

Exclusion criteria for both groups were: Tears >1/3 of the vertical height of the supraspinatus tendon, since the stress may be increased on the intact tendon part, and calcifications >2 mm (length) due to acoustic shadowing. Further exclusion criteria were: previous comorbidities (potentially harmful to the tendon) such as; past/present shoulder fractures, surgery and luxation, known neuromuscular disease, rheumatoid arthritis, cancer, fibromyalgia, spondyloarthropathy and psychiatric disorders. Pregnancy and inability to read and understand Danish were also exclusion criteria.

SEL image capturing and measurement

Apparatus

All measurements were performed with the Logiq S7 using a 15 MHz linear probe (GE Healthcare, Milwaukee, USA). Manufacturer recommendations for musculoskeletal SEL of the shoulder were used, including a transducer movement rate of approximately 120 cycles/min, axial smoothing of 2, lateral smoothing of 3, frequency of 10 and a soft/hard compression of 5.

Patient placement

The SEL was obtained with the patient sitting in an erect posture with the arm internally rotated, elbow flexed to 90° and with the dorsal side of the hand placed over the sacrum, as previously used.[15] The probe was placed on the anterior aspect of the acromion in the coronal plane and the images were obtained just laterally to the anterior-lateral corner of the acromion in the longitudinal plane of the supraspinatus tendon (Figure 1).

Image capturing

An image window depth of at least three times the tendon size and an image width covering about three-quarters of the screen were used as recommended for longitudinal SEL.[22] The tissue was compressed approximately 2-5 mm,[9] and a software incorporated quality control (expressed as 1-5 green bars being displayed, with 5 bars being the most acceptable) was used to evaluate the recommended compression size.

For each assessment method (with or without GEL covering the transducer, (Sonokit (Proxon), thickness: 10 mm, length: 70 mm, elastic modulus: 226 kPa; Sonogel Vertriebs GmbH, Germany)), three sessions of 20 seconds were obtained.

Image measurements

Tendon characteristics were evaluated quantitatively and qualitatively.

Quantitatively, ROIs on the SEL images were drawn over the target area (supraspinatus tendon) and the exact raw strain value (RAW) (0-6; 6 being the hardest tissue) was calculated.

Further, an adjacent reference region (normal tissue, experiencing the same stress as the target region) was drawn. From these two variables, the strain ratio (0-60; 60 being the hardest tissue) was calculated.[23] Two different reference tissues were used: a 1 mm circle region in a soft part of the deltoid/bursal area for the DELT measurement (Figure 2), and a 5 mm circle region in GEL covering the transducer, for the GEL measurement (Figure 3).[13 24] GEL was used as a more homogeneous reference tissue.

For the raw values alone and the strain ratios (strain raw value of supraspinatus tendon (A) to strain raw value of reference tissue (B) = A/B), a mean of the three measured areas of the supraspinatus tendon was calculated including data from 5-15 seconds of the 20-second cycle as recommended by the manufacturer.

Quantitative measurements were based on examination of three entire cine-loops (10 seconds) rather than on single static images,[25] in order to minimise intra-observer variation and avoid transient temporal fluctuations. Only sequences with the highest image quality (with green bars on the quality assessment) were used as recommended by the manufacturer.

Due to difficulties in defining the most lateral part of the supraspinatus insertion on the humeral head, a 6.5 mm chord was drawn (in the lateral direction) from the medial part of the insertion on the humeral head to the lateral part of the tuberculum major. The average length of the supraspinatus insertion has previously been estimated to be 6.5 mm.[26] This fix point (end of the line at the lateral part of the insertion) was used to draw a 23 mm (7.7 x 3) horizontal line in the medial direction (which has been estimated to be the average length of the tendon),[27] ensuring agreement of measurement area (Figure 4).

Caudal borders for image measurements were bony surfaces (of the humeral head), while cranial borders were the transition zone between the superior surface of the supraspinatus tendon and the inferior surface of the deltoid muscle and bursa. The bursa area was used as reference tissue (red colour in the SEL image). If no red border was seen, an estimated border followed the superior surface of the tendon.

Qualitatively, the images were rated using a colour scale, from 1-4, according to the following: Type 1, < 10% colour other than blue (indicating a high tissue stiffness); Type 2, 11-25% colour other than blue; Type 3, 26-50% colour other than blue; or Type 4, > 50% colour other than blue (indicating a low tissue stiffness). Furthermore, the number of yellow/red lesions was graded as follows: 0=no lesions; 1=one lesion; 2=two lesions; and 3=more than two lesions.[14] The qualitative classifications were all performed on the first high quality image recorded closest to 10 seconds into the first cycle.

Development of SEL method

Based on previous studies,[14 16 25 28-31] a protocol with standardised procedures was developed for obtaining SEL characteristics in the Supraspinatus tendon and tested in Phase One on 10 participants.

Based on the results from these participants, adjustments to the type of reference tissue and colour scale criteria were performed. One adjustment involved the replacement of the subcutaneous fat with DELT as reference tissue, since the subcutaneous fat layer in some participants was too thin to measure. Another adjustment was to base the colour scale on the

percentage of colour (replacement of estimation of the most pronounced colour), which made it possible to rate a blue (hard) tendon as softer if it appeared with yellow/red lesions.

In Phase Two, the adjusted protocol was applied to 20 new participants. Overall agreement in Phase Two corresponded to >80% with blinded raters. Thereafter, Phase Three (n=40) was initiated based on the final protocol, as described above.

Questionnaires

Participants completed questionnaires, including the DASH (Disabilities of the Arm, Shoulder and Hand) for investigating disability of the upper extremities (0-100; 100 being most disabled),[32] The VAS (Visual Analogue Scale) for assessing pain level (0-100; 100 being the most painful),[33] the EQ-5D-3L for measuring health-related quality of life, and the EQ-VAS[34] also for health-related quality of life (0-100; 100 being best imaginable health status). Demographic data included information on age, sex, and BMI.

Statistics

SEL data were found to be normally distributed on a histogram. For continuous data (mean of the three measured areas of the supraspinatus tendon) the ICC (model 2.3, absolute agreement, 2-way random, single measures) was calculated to determine intra-rater and inter-rater reliability.[35] The ICC, with 95% confidence intervals (95% C.I.), was calculated for RAW, and for strain ratios (DELT and GEL). The ICCs were interpreted as <0.40=poor, 0.40-0.59=fair, 0.60-0.74=good and ≥ 0.75=excellent reliability.[36]

Paired student t-tests were completed for statistical comparison of ratings (Rater 1 (first time) vs. Rater 1 (second time), and Rater 1 vs. Rater 2) using a significance level of 0.05. Bland and Altman plots with 95% limits of agreement (LOA) were calculated to evaluate systematic differences,[37] with the 95% LOA calculated as the mean difference ± 1.96 x standard deviation of the difference (SD).[38]

Standard Error of Measurements (SEM) was calculated using the formula: SEM = SDmean difference/√2,[39] and Minimal Detectable Change 95% ($MDC_{individual}$) was calculated using the formula: MDC =SEM x √2 x 1.96[40] and the relative MDC was calculated as a percentage of the average SEL values of Rater 1 and Rater 2.

For ordinal data, linear weighted Cohen's κ (LWk) was used to calculate reliability with 95% CIs for colour ratings and number of lesions. LWk was interpreted as <0=poor agreement, 0.01-0.20=slight agreement, 0.21-0.40=fair agreement, 0.41-0.60=moderate agreement, 0.61-0.80=substantial agreement and >0.81=almost perfect agreement.[41]

Statistical analyses were performed in SPSS, version 25.0 (SPSS Inc., Chicago, IL).

RESULTS

Demographics varied between patients and healthy participants, on most parameters. Due to the sampling method, the patients were expected to be older, with a higher BMI, more pain and disability, and reduced quality of life compared with healthy participants (Table 1).

Table 1
Demographics (mean; SD (frequencies and percentages for dominant arm and EQ-5D)) of patients and healthy participants from the study phase (n = 40).

Variable	**Patients (n = 20)**	**Healthy participants (n = 20)**
Age (years)	47.85 (7.63)	25.70 (6.10)
Females (%)	14 (70)	11 (55)
BMI	30.49 (6.57)	24.94 (2.45)
Dominant arm	12 (60%)	10 (50%)
VAS rest (0-100)	24.60 (20.96)	0 (0.00)
VAS activity (0-100)	53.35 (16.62)	0 (0.00)
VAS sleep (0-100)	49.60 (18.26)	0 (0.00)
VAS maximum (0-100)	78.05 (11.62)	0 (0.00)
DASH (0-100)	34.75 (17.48)	1.71 (4.21)
EQ-VAS (0-100)	53.24 (38.31)	72.90 (36.40)
EQ-5D-3L	**Frequencies**	**Frequencies**
Mobility problems	0 (0%)	0 (0%)
Self-care problems	10 (50%)	1 (5%)
Usual activities problems	1 (5%)	0 (0%)
Pain/discomfort problems	20 (100 %)	1 (5%)
Anxiety/depression problems	5 (20%)	1 (5%)

Abbreviations: SD, Standard Deviation; BMI, Body Mass Index; VAS, Visual Analogue Scale; DASH, Disability of the Arm, Shoulder and Hand Questionnaire; EQ-5D-3L, Quality of Life by dimension; EQ-VAS, Quality of Life

The paired t-test showed statistical differences in inter-rater measurements for RAW and in intra-rater measurements for GEL (Table 2).

Table 2

Reliability of strain elastography in the Supraspinatus tendon using respectively the deltoid muscle (reference) and a gel pad (reference) and raw data from the study phase (n = 40).

Continuous scale	Mean (SD)	Mean (SD)	Mean Difference (SD)	P	LOA	SEM	MDC (%)	ICC (95% C.I.)
Mean across tendon								
Intra-rater	**Rater 1**	**Rater 1**						
RAW	4.40 (0.55)	4.36 (0.55)	0.04 (0.14)	0.09	-0.24; 0.33	0.10	0.28 (6.36)	0.97 (0.93-0.98)
DELT (ratio)	14.33 (3.07)	14.23 (3.16)	0.10 (1.48)	0.66	-2.81; 3.01	1.05	2.91 (20.37)	0.89 (0.80-0.94)
GEL (ratio)	2.87 (1.28)	2.59 (1.01)	0.28 (0.82)	0.04*	-1.33; 1.89	0.58	1.61 (58.82)	0.73 (0.54-0.85)
Inter-rater	**Rater 1**	**Rater 2**						
RAW	4.40 (0.55)	4.53 (0.61)	-0.13 (0.24)	0.00*	-0.60; 0.34	0.17	0.47 (10.53)	0.89 (0.75-0.95)
DELT (ratio)	14.33 (3.07)	14.56 (2.60)	-0.22 (1.88)	0.46	-3.91; 3.46	1.33	3.69 (25.51)	0.78 (0.63-0.88)
GEL (ratio)	2.87 (1.28)	2.73 (1.00)	0.14 (0.89)	0.32	-1.61; 1.90	0.63	1.75 (62.63)	0.70 (0.50-0.83)

Abbreviations: SD, Standard Deviation; LOA, Limits of Agreement; SEM, Standard Error of Mean; MDC, Minimal Detectable Change; ICC (95% C.I.), Intra-class Correlation Coefficient with 95% confidence intervals; DELT, Reference area in the Deltoid Muscle; GEL, Reference area in the Gel Pad; RAW, Raw Elastography Data; *, significant difference ($p \leq 0.05$) between measurements.

None of the Bland and Altman plots showed funnel effects (increasing difference with increasing mean size) (Figure 5).

For intra-rater and inter-rater reliability, there was 'excellent agreement' using RAW and DELT (ICC: intra-rater: 0.97, inter-rater: 0.89), (ICC: intra-rater: 0.89, inter-rater: 0.78) respectively and the reliability was 'good' when using GEL (ICC: intra-rater: 0.73, inter-rater: 0.70) (Table 2). The same pattern of results was also seen for each of the three measured areas of the Supraspinatus tendon (medial, middle and lateral parts) (not shown in tables).

For the intra-rater reliability, the relative MDC was smallest for RAW (6.36%), larger for DELT (20.37%) and largest for GEL (58.82%). For inter-rater reliability, the same pattern was seen for MDC, with a minimum of 10.53%, 25.51% and 62.63%, for RAW, DELT and GEL, respectively.

The reliability of using the colour scale (performed without GEL), kappa (LWk) was very similar, with intra-rater reliability ranging from LWk: 0.76-0.79, representing 'substantial agreement', and inter-rater agreement ranging from LWk: 0.71-0.81 representing 'substantial to almost perfect agreement'.

For the number of yellow/red lesions, LWk was generally highest for intra-rater ranging from 0.40-0.82 representing 'fair to almost perfect agreement', while for inter-rater reliability LWk ranged from 0.24-0.67 representing 'fair to substantial agreement' (Table 3).

Table 3

Reliability of strain elastography in the Supraspinatus tendon using data based on colours, respectively a colour scale and counting the number of lesions from the study phase (n=40)

Ordinal scale	**Intra-rater**		**Inter-rater**	
	Total Agreement (%)	LWκ (95% C.I.)	Total Agreement (%)	LWκ (95% C.I.)
Colourscale				
Medial tendon 1/3	80	0.76 (0.64-0.89)	80	0.77 (0.63-0.91)
Middle tendon 1/3	82.5	0.77 (0.65-0.88)	76	0.71 (0.58-0.83)
Lateral tendon 1/3	85	0.79 (0.65-0.94)	87.5	0.81 (0.65-0.97)
Lesions (no.)				
Medial tendon 1/3	87.5	0.82 (0.69-0.95)	82.5	0.67 (0.53-0.80)
Middle tendon 1/3	90	0.75 (0.51-1.00)	87.5	0.63 (0.30-0.96)
Lateral tendon 1/3	87.5	0.40 (-0.16-0.97)	90	0.24 (-0.14-0.62)

Abbreviations: LWκ (95% C.I.), Linear Weighted κ with 95% confidence intervals

For the medial and middle tendon part of the supraspinatus, LWk was 'substantial to almost perfect' for intra-rater (LWk 0.75-0.82) and 'substantial' for inter-rater (LWk 0.63-0.67) reliability, while for the lateral tendon part, LWk was low, corresponding to only 0.40 and 0.24 for intra-rater and inter-rater reliability, respectively, representing 'fair agreement'.

DISCUSSION

The reliability of the results from SEL was 'excellent' when using the raw data and DELT as reference tissue. When using GEL as reference tissue, the reliability of the results was 'good'. Furthermore, the relative MDC (% of mean) was smallest for RAW and largest for GEL.

When using the colour scale grading, LWk represented 'substantial agreement' to 'substantial to almost perfect agreement'. For the number of yellow/red lesions, LWK was highest for intra-rater reliabiliy, in the medial and middle tendon parts, LWk was 'substantial to almost perfect', while for the lateral tendon part LWk was low, corresponding to only 'fair' agreement'.

Strain Ratios and Raw data

The ROIs with raw SEL data (no reference tissue) resulted in the highest reliability in intra-rater as well as inter-rater reliability. Although there was a significant difference between Rater 1 and Rater 2 for RAW, this difference was less than the MDC and is ascribed to measurement error.

To our knowledge, no study has previously presented raw SEL data, which precludes comparison with other studies. Using raw data has limitations, since there is, in comparison with using ratios, no possibilities to adjust for different transducer pressures. On the other hand, the advantage is that it gives a more quantitative estimate than when using visual inspections, as for example in the colour scale, or with uncertainties due to selection of reference tissue.

In the current study, only one rater captured the images (using almost the same pressure) and two raters traced and measured the areas. If more than one rater were to capture the images, a lower reliability may be expected, due to potentially different transducer pressure.

Different reference tissues in SEL measurements of the musculoskeletal area make it difficult to compare data across studies. The reference tissues previously used for calculating strain ratios in the achilles and supraspinatus tendon areas have included bone,[15] fat,[13 25 31 42 43] and GEL/acoustic coupler.[13 16 44] Using bone as a reference value has limitations, since ultrasound cannot penetrate bone material, meaning that data coming from this region are artefacts. In addition, strain ratios may be based on different equations in different studies, placing the ROIs of the tendon in the denominator,[13 45] or, as in the currrent study, in the numerator.[25 31] Due to thin subcutaneous fat tissue area in some participants, the reference tissue was replaced, after Phase One, with DELT and GEL (artificial fat tissue).
When using DELT as the reference tissue, reliability was found to be excellent (both intra-rater and inter-rater). Muscle tissue increases stiffness significantly after exercise and muscle contractions,[16 24] and therefore, limitations are recognised when using this tissue as reference when for example, investigating tendon tissue response to muscle contractions. However, no previous study of SEL on the supraspinatus tendon has used muscle tissue as a reference for the calculated strain ratios.

The current study found a high MDC for GEL (large measurement error) but a good reliability (ICC 0.70-0.73), in line with previous studies. One study, using the Kager's fat pad (fat deposit within the Kager triangle anterior to the achilles tendon) as reference tissue in the

achilles tendon, also found good to excellent intra-rater and inter-tester reliability (ICC: 0.51-0.78) in the longitudinal plane; however, with much lower reliability (ICC: 0.41-0.45) in the axial plane.[25] Unfortunately, MDC was not reported. The ICC and MDC in the current study are reported as single measures, as it is of relevance to clinicians. None of the previous studies[16 44] have described whether ICC is reported as a single measure or (group) average measure, however, it is anticipated that it is a group average measure. In the current study, the ICC will increase about 10% and the MDC will decrease up to 80% when reported as average measures. This could explain some of the differences in results between the current and previous studies.

In the current study, the data using GEL are also in line with a study using an acoustic coupler as reference tissue in the supraspinatus tendon,[16] where excellent intra-rater reliability was shown, as was also confirmed for the achilles tendon.[44] An acoustic coupler is similar to GEL and may be acceptable, but ideally, the reference area should be the same depth as the ROI (in this case, the supraspinatus tendon). As GEL is not located close to the tendon, the ROIs using GEL will not be subjected to the same amount of tissue pressure as the tendon, which may affect reliability and validity.

The current study found the lowest reliability (however, it was still graded 'good') when using GEL which may be caused by difficulties locating the footprint of the tendon, due to lower image quality, because of increased depth (through 10 mm of GEL).

A statistically significant intra-rater difference was also seen, but as this difference was below the MDC, it can be ascribed to measurement error.

Colour Ratings

The current study found a high reliability ('substantial' to 'almost perfect') when grading the ROIs of SEL images according to a 4-level colour scale, where blue tissue indicated hard tissue and red tissue indicated soft tissue. This is in line with a previous study on the achilles tendon using colour scales of 5 levels (1=blue (hardest tissue), 2=light blue, 3=green, 4=yellow, 5=red (softest tissue)), where good to excellent inter-rater (κ = 0.897) and intra-rater (κ = 1) reliability was found.[30]

Alternative types of colour scales have been used in the musculoskeletal area, primarily for the achilles tendon, with categorisation of the achilles tendon into a two-level category scale (green/blue vs. red),[25] and a three-level category scale (blue/yellow/red).[22 29] The current study has used the same colours for defining hard tissue (blue), and for soft tissue (red), but with a four-level category scale, and with more precise criteria for the different levels which were found necessary due to the mixture of colours in the supraspinatus tendon observed in Phase One.

In the current study, LWk in the medial and the middle parts of the tendon was 'substantial to almost perfect' when counting the number of yellow/red lesions in both intra-rater and inter-rater reliability. Even though the total agreement was also high in the lateral third of the tendon, LWk was relatively low compared with the medial and the middle parts of the tendon. The reason may be due to the low presence of lesions in the lateral part, corresponding to only 10 % of the participants presenting with lesions in the lateral 1/3 of the tendon, which can lead to the 'Kappa Paradox'.[46] The Kappa Paradox means (in a 2x2 table) that an imbalance between presence and absence in overall agreement, and between

disagreement (bias) can drastically lower the kappa, why prevalence and bias-adjusted kappa (PABAK) in dichotomous scales is recommended.[47] A PABAK on the current number of lesions dichotomised into no lesion /lesion(s) will increase the kappa by up to 70%. Another explanation is that since the pressure is placed vertically on the medial part of the tendon, the lateral part will have experienced a smaller degree of stress.

The high reliability of the number of focal lesions (LwK intra-rater reliability: 87.5-90, LwK inter-rater reliability: 82.5-90) is in line with a previous study on the supraspinatus tendon where an almost prefect reliability (k = 0.83) was found.[14]

The current methods of using colour grading and lesion counts for assessing tendon stiffness in the supraspinatus tendon showed high reliability. These methods are feasible for use in clinical practice as they can be performed quickly, but when more than one clinician is performing the SEL, the method cannot adjust for potentially different tranducer pressures.

Limitations

Manual compression may affect reliability, especially when using raw data, colour grading and counting yellow/red lesions. To partly counteract this, a quality bar was used that provided instant feedback on the uniformity of the transducer pressure. Reference tissues (DELT and GEL) with calculation of strain ratios were also used, thereby making a comparison possible between the methods.

Furthermore, as mentioned, SEL is highly operator-dependent, why the same (and only one) trained operator captured the present images, thereby further decreasing the risk of bias, but this limits the external validity of the results.

In addition, age was not blinded and, as reported earlier, tendons get softer with age.[48] Therefore, the current blinding of examiner results and health status may not have been sufficient.

Strengths

The strength of this study was the design, incorporating a stepwise and standardised procedure for reliability which minimised bias and increased reliability.[17] Furthermore, both patients and healthy participants were enrolled.

To enhance standardisation, all SEL ROIs were measured at a fixed point, just laterally from the anterior-lateral corner of the acromion in the longitudinal plane of the supraspinatus tendon.[20] In addition, the reliability was estimated by using three different quantitative methods (RAW, DELT and GEL), as well as two different qualitative methods (colour and number of yellow/red focal lesions).

CONCLUSION

Intra-rater and inter-rater reliability were excellent using raw values and ratios with the deltoid muscle as reference tissue, and good when using a gel pad as reference tissue. The reliability of colour scale ratings was substantial to almost perfect and the number of lesions was fair to almost perfect.

Although high reliability was found, validity and responsiveness of these elastographic methods needs further investigation.

Contributorship statement

KB, JH, PK, KGI and BJ-K conceived and designed the study protocol. KB and BJ-K procured the project funding. KB, JH, PK, and BJ-K developed and standardised the ultrasound procedure. JH, KB and KGI recruited participants.

KB was the project coordinator and captured the strain elastography images. KGI performed the physical tests. KB and JH rated the images. KB and BJ-K planned and coordinated the statistical analyses. KB performed the statistical analyses. KB drafted the manuscript, and JH, PK, KGI and BJ-K contributed to the manuscript. All authors read and approved the final manuscript. KB is the guarantor.

Competing interests

The authors declare no conflicts of interest.

Funding

This work was supported by the Region of South Denmark (14/38969, 15/24799), The Danish Rheumatism Association (R150-A4296) and the Danish Council of Radiographers (no grant number).

Data sharing statement

No additional data are available.

REFERENCES

1. Luime JJ, Koes BW, Hendriksen IJ, et al. Prevalence and incidence of shoulder pain in the general population; a systematic review. *Scand J Rheumatol* 2004;33(2):73-81.

2. Singh SG, Sumit.; Mohammad, F.; Kumar, S.; Kumar, D.; Kumar, S. Prevalence of shoulder disorders in tertiary care centre. *J Res Med Sci* 2015;3(4):917-20.

3. Nyman PP, K.; Panula, H.; Mälkiä, E. Patients' experience of shoulder problems prior to and following intervention. *Physiother Theory Pract* 2012;28(3):221-31.

4. Virta L, Joranger P, Brox JI, et al. Costs of shoulder pain and resource use in primary health care: a cost-of-illness study in Sweden. *BMC Musculoskelet Disord* 2012;13:17.

5. Teunis T, Lubberts B, Reilly BT, et al. A systematic review and pooled analysis of the prevalence of rotator cuff disease with increasing age. *J Shoulder Elbow Surg* 2014;23(12):1913-21.

6. Klauser AS, Faschingbauer R, Jaschke WR. Is sonoelastography of value in assessing tendons? *Semin Musculoskelet Radiol* 2010;14(3):323-33.

7. Prado-Costa R, Rebelo J, Monteiro-Barroso J, et al. Ultrasound elastography: compression elastography and shear-wave elastography in the assessment of tendon injury. *Insights Imaging* 2018

8. Dietrich CF, Barr RG, Farrokh A, et al. Strain Elastography - How To Do It? *Ultrasound Int Open* 2017;3(4):E137-e49.

9. Garra BS. Elastography: history, principles, and technique comparison. *Abdom Imaging* 2015;40(4):680-97.

10. Chung SW, Kim JY, Yoon JP, et al. Arthroscopic repair of partial-thickness and small full-thickness rotator cuff tears: tendon quality as a prognostic factor for repair integrity. *Am J Sports Med* 2015;43(3):588-96.

11. Bamber J, Cosgrove D, Dietrich CF, et al. EFSUMB guidelines and recommendations on the clinical use of ultrasound elastography. Part 1: Basic principles and technology. *Ultraschall Med* 2013;34(2):169-84.

12. Fusini F, Langella F, Busilacchi A, et al. Real-time sonoelastography: principles and clinical applications in tendon disorders. A systematic review. *Muscles Ligaments Tendons J* 2017;7(3):467-77.

13. Lee SU, Joo SY, Kim SK, et al. Real-time sonoelastography in the diagnosis of rotator cuff tendinopathy. *J Shoulder Elbow Surg* 2016;25(5):723-9.

14. Seo JB, Yoo JS, Ryu JW. Sonoelastography findings of supraspinatus tendon in rotator cuff tendinopathy without tear: comparison with magnetic resonance images and conventional ultrasonography. *J Ultrasound* 2015;18(2):143-9.

15. Tudisco C, Bisicchia S, Stefanini M, et al. Tendon quality in small unilateral supraspinatus tendon tears. Real-time sonoelastography correlates with clinical findings. *Knee Surg Sports Traumatol Arthrosc* 2013

16. Muraki T, Ishikawa H, Morise S, et al. Ultrasound elastography-based assessment of the elasticity of the supraspinatus muscle and tendon during muscle contraction. *J Shoulder Elbow Surg* 2015;24(1):120-6.

17. Patijn J. Reproducibility and Validity Studies of Diagnostic Procedures in Manual/Musculoskeletal Medicine 2004.

18. Kottner J, Audige L, Brorson S, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol* 2011;64(1):96-106.

19. Hougs Kjaer BPM, Ellegaard KPP, Wieland IP, et al. Intra-rater and inter-rater reliability of the standardized ultrasound protocol for assessing subacromial structures. *Physiother Theory Pract* 2017;33(5):398-409.

20. Ingwersen KG, Hjarbæk J, Eshøj H, et al. Ultrasound assessment for grading structural tendon changes in supraspinatus tendinopathy: an inter-rater reliability study. *B M J Open* 2016;6(5)

21. Sein ML, Walton J, Linklater J, et al. Reliability of MRI assessment of supraspinatus tendinopathy. *Br J Sports Med* 2007;41(8):e9.

22. De Zordo T, Fink C, Feuchtner GM, et al. Real-time sonoelastography findings in healthy Achilles tendons. *AJR Am J Roentgenol* 2009;193(2):W134-8.

23. Cosgrove D, Piscaglia F, Bamber J, et al. EFSUMB guidelines and recommendations on the clinical use of ultrasound elastography. Part 2: Clinical applications. *Ultraschall Med* 2013;34(3):238-53.

24. Niitsu M, Michizaki A, Endo A, et al. Muscle hardness measurement by using ultrasound elastography: a feasibility study. *Acta Radiol* 2011;52(1):99-105.

25. Drakonaki EE, Allen GM, Wilson DJ. Real-time ultrasound elastography of the normal Achilles tendon: reproducibility and pattern description. *Clin Radiol* 2009;64(12):1196-202.

26. Lumsdaine W, Smith A, Walker RG, et al. Morphology of the humeral insertion of the supraspinatus and infraspinatus tendons: Application to rotator cuff repair. *Clin Anat* 2015;28(6):767-73.

27. Curtis AS, Burbank KM, Tierney JJ, et al. The insertional footprint of the rotator cuff: an anatomic study. *Arthroscopy* 2006;22(6):609.e1.

28. De Zordo T, Lill SR, Fink C, et al. Real-time sonoelastography of lateral epicondylitis: comparison of findings between patients and healthy volunteers. *AJR Am J Roentgenol* 2009;193(1):180-5.

29. Klauser AS, Miyamoto H, Tamegger M, et al. Achilles tendon assessed with sonoelastography: histologic agreement. *Radiology* 2013;267(3):837-42.

30. Sconfienza LM, Silvestri E, Cimmino MA. Sonoelastography in the evaluation of painful Achilles tendon in amateur athletes. *Clin Exp Rheumatol* 2010;28(3):373-8.

31. Turan A, Tufan A, Mercan R, et al. Real-time sonoelastography of Achilles tendon in patients with ankylosing spondylitis. *Skeletal Radiol* 2013;42(8):1113-8.

32. Kitis A, Celik E, Aslan UB, et al. DASH questionnaire for the analysis of musculoskeletal symptoms in industry workers: a validity and reliability study. *Appl Ergon* 2009;40(2):251-5.

33. Boonstra AM, Schiphorst Preuper HR, Reneman MF, et al. Reliability and validity of the visual analogue scale for disability in patients with chronic musculoskeletal pain. *Int J Rehabil Res* 2008;31(2):165-9.

34. Norman R, Cronin P, Viney R, et al. International comparisons in valuing EQ-5D health states: a review and analysis. *Value Health* 2009;12(8):1194-200.

35. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;86(2):420-8.

36. Cicchetti D. Guidelines, Criteria, and Rules of Thumb for Evaluating Normed and Standardized Assessment Instrument in Psychology. *Psychol Assess* 1994;6(4):284-90.

37. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1(8476):307-10.

38. Ottenbacher KJ, Stull GA. The analysis and interpretation of method comparison studies in rehabilitation research. *Am J Phys Med Rehabil* 1993;72(5):266-71.

39. de Vet HCW, Terwee CB, Knol DL, et al. When to use agreement versus reliability measures. *J Clin Epidemiol* 2006;59(10):1033-9.

40. de Vet HCW, Terwee CB, Knol DL, et al. Measurements in Medicine - a pratical guide. New York, USA: Cambridge University Press 2011.

41. Landis RJK, G.G. The Measurement of Observer Agreement for Categorical Data. *Biometrics* 1977;33(1):159–74.

42. Busilacchi A, Olivieri M, Ulisse S, et al. Real-time sonoelastography as novel follow-up method in Achilles tendon surgery. *Knee Surg Sports Traumatol Arthrosc* 2014

43. Ooi CC, Schneider ME, Malliaras P, et al. Diagnostic performance of axial-strain sonoelastography in confirming clinically diagnosed Achilles tendinopathy: comparison with B-mode ultrasound and color Doppler imaging. *Ultrasound Med Biol* 2015;41(1):15-25.

44. Yamamoto Y, Yamaguchi S, Sasho T, et al. Quantitative Ultrasound Elastography With an Acoustic Coupler for Achilles Tendon Elasticity: Measurement Repeatability and Normative Values. *J Ultrasound Med* 2016;35(1):159-66.

45. Akkaya S, Akkaya N, Agladioglu K, et al. Real-time elastography of patellar tendon in patients with auto-graft bone-tendon-bone anterior cruciate ligament reconstruction. *Archives of Orthopaedic and Trauma Surgery* 2016;136(6):837-42.

46. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol* 1990;43(6):543-9.

47. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *J Clin Epidemiol* 1993;46(5):423-9.

48. Hsiao MY, Chen YC, Lin CY, et al. Reduced patellar tendon elasticity with aging: In vivo assessment by shear wave elastography. Ultrasound Med. Biol. 2015;41(11):2899-905.

Figure legends

Figure 1

With the arm located behind the back and the elbow flexed at 90° and the palm facing posteriorly, the probe was placed just laterally from the anterior-lateral corner of the acromion in the longitudinal plane of the supraspinatus tendon.

Figure 2

Left side: For measuring elastography characteristics of the supraspinatus tendon, the tendon was split into three parts (3 x 7.7mm), illustrated by the areas of blue, red and green colours. This division was based on a line from the lateral tendon insertion (greater tubercle) to the medial tendon insertion, corresponding to 6.5 mm, and from there a line of 23 mm to the medial tendon with the end point (medial part) being perpendicular to the superior surface of the tendon.

The yellow circle in the soft part of the deltoid muscle is used as the reference tissue.

Right side: The three measurement areas (ROIs) and one reference area with elastography characteristics (raw data) during the time of measurement.

Figure 3

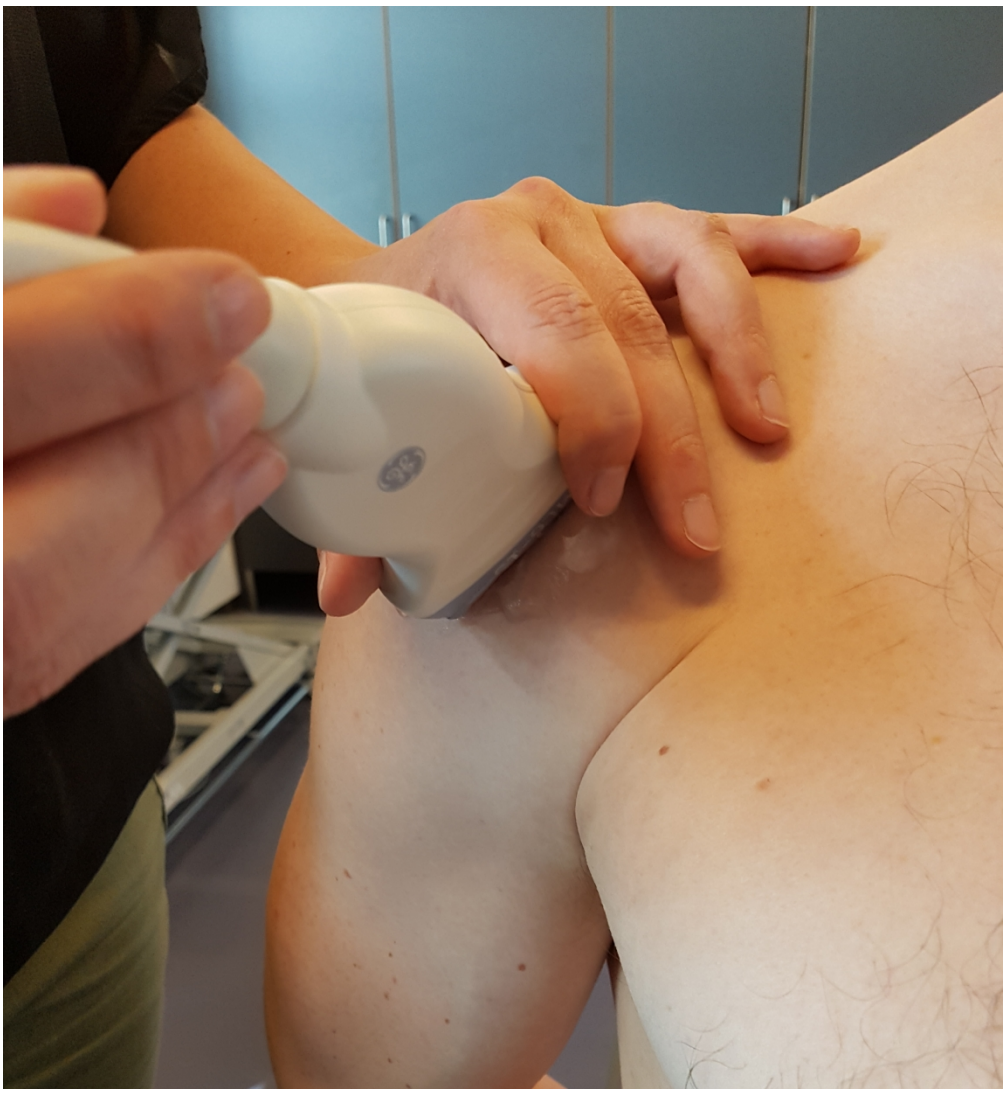
Left side: Gel pad (mounted on the transducer with a condom used as the other reference area, yellow circle). Right side: The corresponding elastography image/measurements.

Figure 4

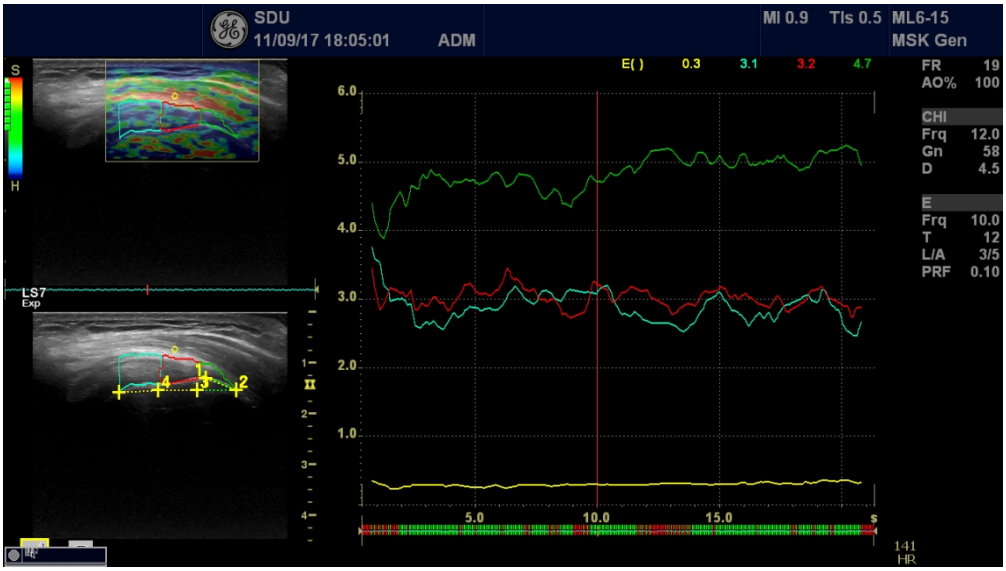
A 6.5 mm chord was drawn (in a lateral direction) from the medial part of the insertion on the humeral head to the lateral part of the tuberculum major. This fixed point (end of the line at the lateral part of the insertion) was used to draw a 23 mm (7.7 x 3) horizontal line in a medial direction ensuring agreement of measuring area.

Figure 5

Bland-Altman plots with 95% Limits of Agreement for measurements, using respectively raw data, the deltoid muscle and a gel pad as reference areas. Values are based on the mean of the three measured parts across the supraspinatus tendon.

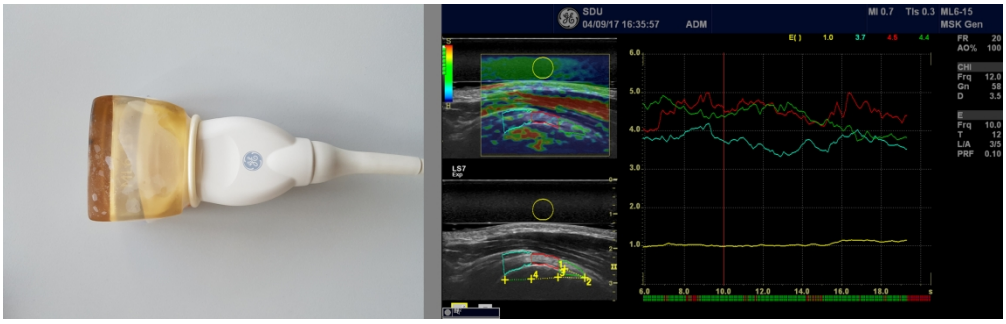


With the arm located behind the back and the elbow flexed 90° with the palm facing towards the posterior direction, the probe is placed just laterally from the anterior-lateral corner of the acromion in the longitudinal plane of the supraspinatus tendon.

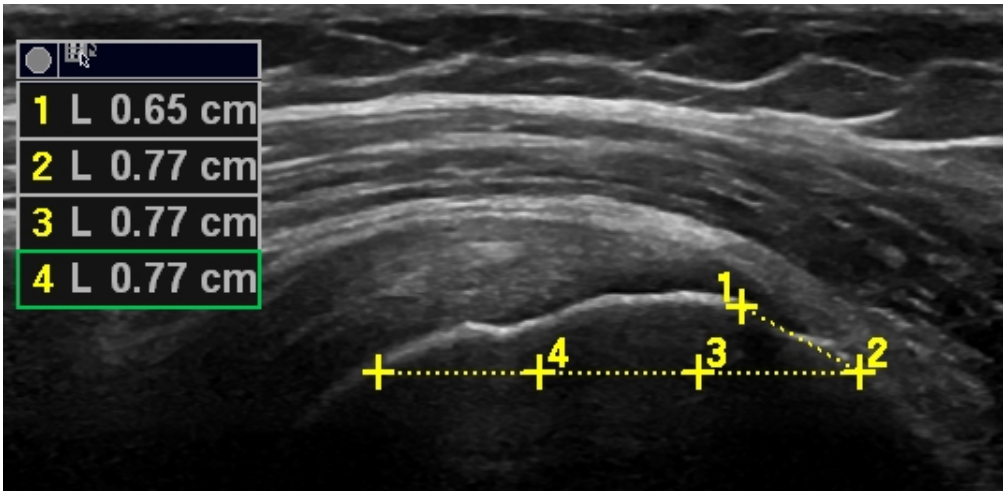


Left side: For measuring elastography characteristics of the supraspinatus tendon, the tendon was split into 3 parts (3x7.7mm), illustrated by the areas of blue, red and green colors. This division was based on a line from the lateral tendon insertion (tuberculum majus) to the medial tendon insertion part, corresponding to 6.5 mm, and from there a line of 23 mm to the medial tendon part with the end point (medial part) being perpendicular to the superior surface of the tendon.
The yellow circle in the soft part of the deltoid muscle is used as a reference.
Right side: The three measurements areas (ROI's) and one reference area with elastography characteristics (raw data) during the time of measurement.

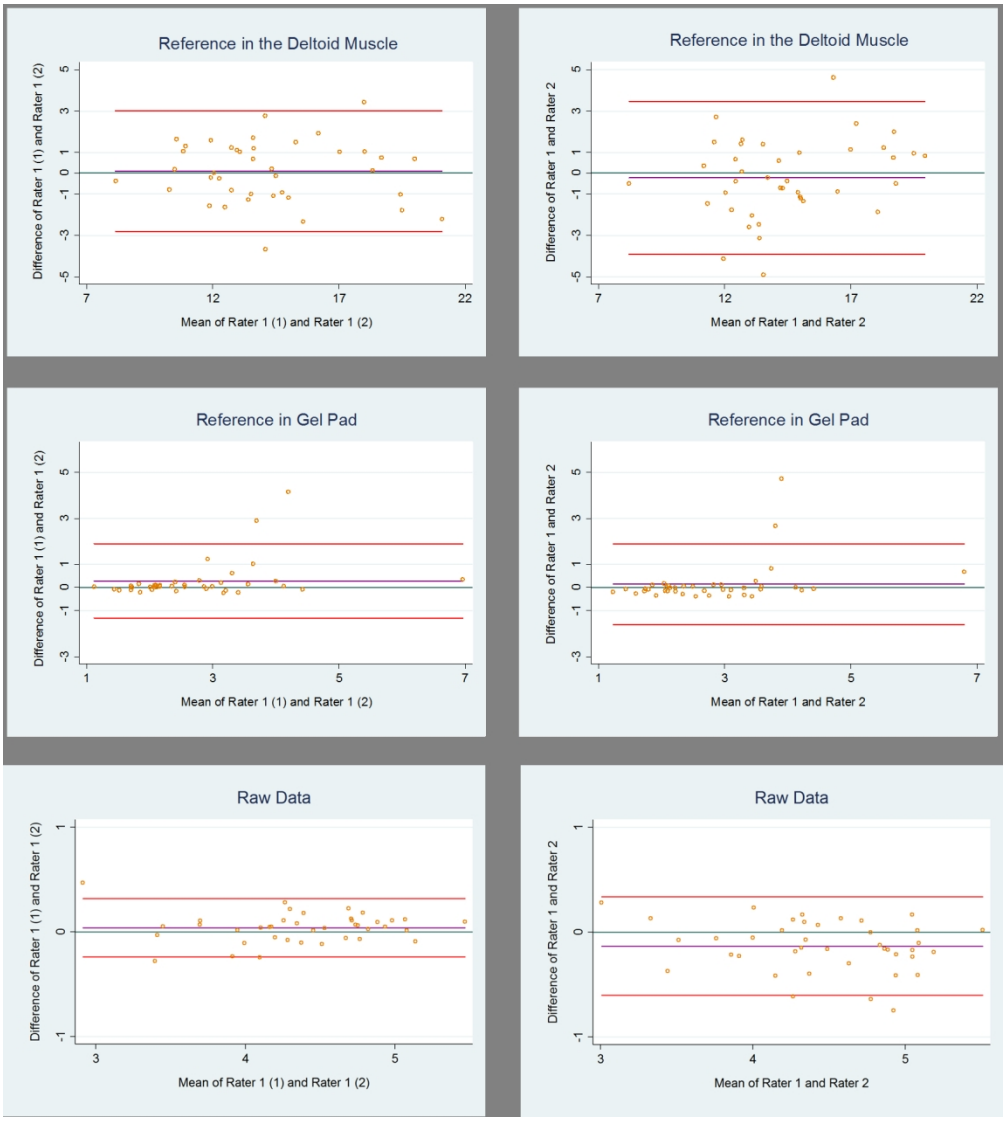
451x254mm (72 x 72 DPI)



Left side: Gel pad (mounted on the transducer with a condom used as the other reference area, yellow circle). Right side: The corresponding elastography image/measurements.



A 6.5 mm chord was drawn (in the lateral direction) from the medial part of the insertion on the humeral head to the lateral part of the tuberculum major. This fix point (end of the line at the lateral part of the insertion) was used to draw a 23 mm (7.7 x 3) horizontal line in the medial direction ensuring agreement of measuring area.



Bland-Altman plots with 95% Limits of Agreement for measurements, using respectively raw data, the deltoid muscle and a gel pad as reference areas. Values are based on the mean of the three measured parts across the supraspinatus tendon.

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract **(p. 1, l. 1)**
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found **(p. 2, l. 1)**
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported **(p. 4, l. 1)**
Objectives	3	State specific objectives, including any prespecified hypotheses **(p. 6, l. 11)**
Methods		
Study design	4	Present key elements of study design early in the paper **(p. 7, l. 2)**
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection **(p. 7, l. 14)**
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up **(p. 8, l. 1)**
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable **(p. 10, l. 19)**
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group **(p. 10, l. 19)**
Bias	9	Describe any efforts to address potential sources of bias **(p. 6, l. 1 & (p. 11, l. 16)**
Study size	10	Explain how the study size was arrived at **(p. 4, l. 5)**
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why **(p. 13, l. 16)**
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding
		(*b*) Describe any methods used to examine subgroups and interactions
		(*c*) Explain how missing data were addressed
		(*d*) If applicable, explain how loss to follow-up was addressed
		(*e*) Describe any sensitivity analyses
Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed
		(b) Give reasons for non-participation at each stage
		(c) Consider use of a flow diagram
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders **(p. 15, l. 5)**
		(b) Indicate number of participants with missing data for each variable of interest
		(c) Summarise follow-up time (eg, average and total amount)
Outcome data	15*	Report numbers of outcome events or summary measures over time
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included
		(*b*) Report category boundaries when continuous variables were categorized

		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses
Discussion		
Key results	18	Summarise key results with reference to study objectives **(p. 20, l. 1)**
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias **(p. 24, l. 16)**
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence **(p. 20, l. 18)**
Generalisability	21	Discuss the generalisability (external validity) of the study results
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based **(p. 27, l. 1)**

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.