Subacromial space measured by ultrasound imaging in asymptomatic subjects and patients with subacromial shoulder pain : an inter-rater reliability study

Reference:
Cavaggion Claudia, Navarro Ledesma Santiago, Luque-Suarez Alejandro, Juul-Kristensen Birgit, Voogt Lennard, Struyf Filip.- Subacromial space measured by ultrasound imaging in asymptomatic subjects and patients with subacromial shoulder pain : an inter-rater reliability study
Physiotherapy: theory and practice - ISSN 1532-5040 - (2022), p. 1-12
Full text (Publisher's DOI): https://doi.org/10.1080/09593985.2022.2072251
To cite this reference: https://hdl.handle.net/10067/1881150151162165141

Subacromial space measured by ultrasound imaging in asymptomatic subjects and patients with subacromial shoulder pain: an inter-rater reliability study

Claudia Cavaggion[a], Santiago Navarro-Ledesma[b], Alejandro Luque-Suarez[c,d], Birgit Juul-Kristensen[e], Lennard Voogt[f,g], Filip Struyf[a*]

[a] *Research Group MOVANT, Department of Rehabilitation Sciences and Physiotherapy (REVAKI), University of Antwerp, Belgium*

[b] *Department of Physiotherapy, University of Granada, Spain*

[c] *Department of Physiotherapy, University of Malaga, Spain*

[d] *Instituto de Investigacion Biomedica de Malaga (IBIMA), Spain*

[e] *Department of Sports Science and Clinical Biomechanics, University of Southern Denmark, Denmark*

[f] *Department of Physical Therapy Studies and Research Centre for Health Care Innovations, Rotterdam University of Applied Sciences, Rotterdam, The Netherlands*

[g] *Pain in Motion Research Group (PAIN), Department of Physiotherapy, Human Physiology and Anatomy, Faculty of Physical Education and Physiotherapy, Vrije Universiteit Brussel, Belgium*

*corresponding author: Professor Filip Struyf

Email: filip.struyf@uantwerpen.be

Address: Department of Rehabilitation Sciences and Physiotherapy, Faculty of Medicine and Health Sciences, University of Antwerp, Campus Drie Eiken - Room D.S.032, Universiteitsplein 1, Wilrijk 2610, Belgium

Subacromial space measured by ultrasound imaging in asymptomatic subjects and patients with subacromial shoulder pain: an inter-rater reliability study

Background: Acromiohumeral distance (AHD) and supraspinatus tendon thickness (STT) measured by Ultrasound (US) can be combined in the occupation ratio (OR). Inter-rater reliability on these subacromial measures depends on the US experience of raters and on the subject status, differing between asymptomatic or patients with subacromial shoulder pain (SSP).

Objective: to evaluate inter-rater reliability between two raters with different US experience (experienced examiner and novice examiner).

Methods: 20 asymptomatic subjects (controls) and 21 patients with SSP were examined on one shoulder. Inter-rater reliability was evaluated with Intraclass Correlation Coefficient (ICC).

Results: ICC for controls was good for AHD at rest and at 60° (0.76-0.77), moderate for STT and AHD at 60° with weights (0.53-0.72), while OR was poorly reliable (below 0.44). ICC for SSP was moderate for AHD at rest and at 60°, STT, OR at rest and at 60° (0.52-0.74) and poor for AHD at 60° with weights and OR at 60° with weights (0.33-0.36). Bland-Altman plots showed systematic bias.

Conclusion: inter-rater reliability varied largely from poor to good between two examiners with different US experience. Clinicians might use the US as additional tool for detecting the subacromial structures, but a structured training including also symptomatic subjects is suggested. Researchers may further investigate the OR in matched case-control studies and an overall agreement phase is recommended before starting the inter-rater reliability phase.

KEYWORDS: acromiohumeral distance; supraspinatus tendon; reproducibility of results; shoulder pain; ultrasonography

INTRODUCTION

Shoulder pain creates functional disability with major socio-economic implications (Kulkarni et al, 2015; Virta, Joranger, Brox and Eriksson, 2012), with subacromial shoulder pain (SSP) being the most common diagnosis in a Swedish cost-of-illness study (Virta, Joranger, Brox and Eriksson, 2012). SSP is generally located in the superior-lateral side of the shoulder and different terms such as subacromial impingement, rotator cuff tendinopathy, supraspinatus tendinitis or bursitis are used in the literature, often referring to a spectrum of the same condition (Kulkarni et al, 2015).

Ultrasound (US) imaging is a non-invasive and non-ionizing medical imaging modality which can clarify the pathomechanics of the SSP, particularly regarding the contribution of the narrowing of the subacromial space, identified as acromiohumeral distance (AHD) on the US image (McCreesh, Crotty and Lewis, 2015). A reduction of the subacromial space is hypothesized as result of external compression (or impingement) of the rotator cuff above the humeral head and under the anterior part of the acromion and the coraco-acromial ligament (Neer, 1972). The AHD has been studied to detect changes of the subacromial space using different rehabilitation programs and to investigate correlations with pain and function (Desmeules et al, 2004; Savoie et al, 2015). Besides the size of the subacromial space, the supraspinatus tendon thickness (STT) has been extensively investigated, as it is the mostly affected structure in SSP (McCreesh, Crotty and Lewis, 2015). Moreover, STT and AHD can be combined in the occupation ratio (OR), which represents the proportion of the subacromial space occupied by the supraspinatus tendon (Michener et al, 2015). The OR can help to explain the possible relationship between extrinsic (tendon compression, represented by AHD) and intrinsic (tendon swelling, indicated by increased STT) mechanisms in SSP (McCreesh, Purtill, Donnelly and Lewis, 2017) and it warrants further attention.

It is important to test the reliability of these US measure before implementing them in clinical practice. US imaging has shown excellent intra-rater reliability (ICC above 0.90) on both STT (Bağcier, Geler Külcü, Yorulmaz and Altunok, 2020; Hougs Kjaer et al, 2017; McCreesh, Anjum, Crotty and Lewis, 2016) and AHD (Bağcier, Geler Külcü, Yorulmaz and Altunok, 2020; Hougs Kjaer et al, 2017; McCreesh, Anjum, Crotty and Lewis, 2016; Pijls et al, 2010) in asymptomatic and symptomatic subjects. However, inter-rater reliability is less frequently investigated, reporting conflicting results. AHD measured at rest showed from good (ICC=0.78) (Desmeules et al, 2004), to excellent (ICC=0.96) (Hougs Kjaer et al, 2017) inter-rater reliability in asymptomatic patients. In contrast, inter-rater reliability of AHD at rest varied more in symptomatic subjects, ranging from moderate (ICC=0.70) (Pijls et al, 2010) to excellent (ICC above 0.90) (Bağcier, Geler Külcü, Yorulmaz and Altunok, 2020; Hougs Kjaer et al, 2017; McCreesh, Anjum, Crotty and Lewis, 2016). Overall, inter rater-reliability in AHD at rest was higher when there was no reported difference in US experience between raters (Bağcier, Geler Külcü, Yorulmaz and Altunok, 2020; Desmeules et al, 2004; McCreesh, Anjum, Crotty and Lewis, 2016) or, if a difference was present, a preliminary agreement phase was conducted (Hougs Kjaer et al, 2017). When AHD was measured at 60° of active abduction, differences in inter-rater reliability increased, ranging from moderate (ICC=0.68) (Mackenzie, Bdaiwi, Herrington and Cools, 2016) to excellent (ICC above 0.90) (Desmeules et al, 2004) in asymptomatic subjects. When AHD at 60° was measured in patients with subacromial impingement by raters with different US experiences (i.e. experienced radiologist and novice orthopedic resident), the ICC decreased down to 0.64 (Pijls et al, 2010).

Concerning STT, excellent inter-rater reliability (ICC above 0.90) was reported in patients with shoulder pain (Bağcier, Geler Külcü, Yorulmaz and Altunok, 2020; Hougs Kjaer et al, 2017; McCreesh, Anjum, Crotty and Lewis, 2016), while only one study reported inter-rater reliability specifically in asymptomatic subjects, ranging from good to excellent (ICC 0.82-0.93) (Hougs Kjaer et al, 2017). Similarly to the inter-rater reliability in AHD at rest, ICC values in STT were higher between raters when there was no reported difference in US experience (Bağcier, Geler Külcü, Yorulmaz and Altunok, 2020; McCreesh, Anjum, Crotty and Lewis, 2016) or when a preliminary agreement phase was performed (Hougs Kjaer et al, 2017).

The OR is a relatively new parameter and its reliability depends on the measures of STT and AHD. Only one study evaluated the reliability of OR in symptomatic subjects, reporting good to excellent (ICC=0.88-0.91) intra-reliability and good (ICC=0.79) inter-rater reliability (Bağcier, Geler Külcü, Yorulmaz and Altunok, 2020). As inter-rater reliability studies on AHD at different angles and STT in SSP and asymptomatic subjects are scarce or reporting different levels of ICC, the aim of this study was to assess the inter-rater reliability between an experienced US examiner and a novice. The inter-rater reliability was also investigated for the calculated OR at rest, at 60° in scapular plane elevation (SPE) and at 60° in SPE with weights. To the best of our knowledge, this is the first study assessing the AHD and the OR at 60° of SPE with weights.

METHODS

This study is reported following the Guidelines for Reporting Reliability and Agreement Studies (GRRAS) (Kottner et al, 2011) and it examined the inter-reliability on seven US measures: AHD at rest, AHD at 60° in SPE, AHD at 60° in SPE with weights, STT, OR at rest, OR at 60° in SPE, OR at 60° with weights. Ethical approval was obtained by the Ethics committee of Antwerp University Hospital (ref: B300201837376) and all subjects gave written informed consent.

Training Phase

Prior to the inter-rater reliability study, rater A (novice US examiner) was trained under supervision of rater B (experienced US examiner) during a 3-days research stay in August 2019 on three asymptomatic subjects. Rater A was a physical therapist who had already 3 years of non-continuous research-related US experience, while rater B was a physical therapist with 9 years of clinical and US experience. Different US measures and approaches (examiner and probe positions) were discussed to decide which US measurements to include in the inter-rater reliability study.

Inter-Rater Reliability Phase

In February 2020 three pilot asymptomatic subjects were tested prior to the start of the inter-rater reliability phase to confirm the US measures and relative modalities of image acquisition learned during the training phase. After this pilot testing, the inter-rater reliability phase lasted 2 weeks. The design of the inter-rater reliability study was cross-sectional, and two groups of subjects (SSP and asymptomatic subjects) were included to give a clear representation of AHD, STT and OR in normal and pathologic tendons.

Participants

Patients with SSP were recruited at three different physiotherapy practices in Belgium. They were invited by the treating physiotherapists to enroll in the inter-rater reliability study on US imaging measures. The inclusion criteria were: age between 18-68 years old; unilateral or bilateral shoulder pain for at least 3 months; at least three of the following shoulder impingement tests positive: Neer test, Hawkins-Kennedy test, Jobe test, painful arc test and external rotation resistance test (Michener, Walsworth, Doukas and Murphy, 2009). The exclusion criteria were: corticosteroid injections within the past 6 weeks; cervical or thoracic surgery within the past 10 years; previous surgery or fractures on the tested shoulder; dislocations on the tested shoulder in the past 10 years; recent (within the past 3 months) US examinations confirming full-thickness RC tears or calcifications larger than 5 mm; presence of concurrent competing pathologies (shoulder arthritis, neurological disorders, fibromyalgia, shoulder malignancy).

Asymptomatic participants were recruited among university staff and friends by personal communication. They were included if they met the following criteria: no pain on the shoulder examined within the past 3 months, no previous operations on the shoulder examined, and did not test positive on more than one of the impingement tests for SSP. Furthermore, they also completed the same clinical questionnaires as the patients with SSP in order to objectively evaluate the differences between symptomatic and asymptomatic subjects. The dominant side was examined with the US for asymptomatic controls.

Demographic details were collected from all subjects, together with the Shoulder and Pain Disability Index (SPADI) and Numeric Pain Rating Scale (NPRS) at rest, during activity and in the past week (on average). Both scales are deemed psychometrically reliable and valid in patients with shoulder pain (Elvers, Oostendorp and Sierevelt, 2003; Mintken, Glynn and Cleland, 2009; Roach, Budiman-Mak, Songsiridej and Lertratanakul, 1991; Thoomes-de Graaf et al, 2015). Considering 10% of dropout, 21 subjects were required to reach an expected ICC of 0.90 (with a minimum value of 0.70), alpha value of 0.05 and power of 0.80 using 2 raters (Bujang and Baharum, 2017). Since reliability can vary in asymptomatic and symptomatic populations, 42 subjects were required (21 asymptomatic and 21 patients with SSP).

Procedure

Rater A enrolled the subjects and then they filled in the clinical questionnaires. Rater A performed the orthopedic tests, took three images of the same US structure (AHD, STT, and

AHD at 60° with or without weights) and measured the parameters on the screen. When the US examination was completed, rater A saved the images on the US machine. The second examiner (US expert, rater B) repeated the process in the same order directly after rater A. The order of measurement was not randomized and raters were blinded for the results of each other but they were aware of the status (asymptomatic or symptomatic) of the subjects examined. The probe was repositioned at each trial and subjects could move the arm between measurements if they felt pain or tiredness. The means of three trials for all US structures were calculated only after the end of the data collection for all subjects and the OR was calculated afterwards.

<u>US Measurements</u>

The US imaging measurements were conducted with a GE Logiq-V2 with a 4.2-13.0 MHz linear-array transducer (GE Healthcare). A preset for musculoskeletal examination with frequency at 12 MHz and "Coded Harmonic Imaging" was used for the study period. However, the US examiner could change depth, gain or focus as necessary. All the measurements were taken in millimeters.

Acromiohumeral Distance at Rest

The subject was seated on an armless chair with the back straight, both feet flat on the ground, head facing forward and the arm examined resting at the side. The US probe was positioned on the most anterior aspect of the acromion, with the long axis of the transducer in the direction of the scapular plane (Navarro-Ledesma and Luque-Suarez, 2018). AHD was defined as the shortest distance between the most inferior edge of the acromion and the humerus (Desmeules et al, 2004; Navarro-Ledesma and Luque-Suarez, 2018). When the bony landmarks were clearly visualized, the US image was frozen and the measurement was taken directly on the screen (see Figure 1).

Supraspinatus Tendon Thickness

The subject was seated in the modified Crass position, with the palm placed on "back pocket" and the elbow flexed and pointing posteriorly (Martinoli, 2010; Navarro-Ledesma, Struyf, Falla and Luque-Suarez, 2019). The STT was examined only in the short axis (transverse view), as this was the view used by Michener et al. for the calculation of the occupation ratio (Michener et al, 2015). The transducer was placed on the anterior aspect of the shoulder and perpendicular to the supraspinatus tendon (Navarro-Ledesma, Struyf, Falla and Luque-Suarez, 2019). With the tendon in view, the transducer was moved until the long head of the biceps (LHB) was visualized and then the image was frozen (McCreesh, Anjum, Crotty and Lewis, 2016). Since the supraspinatus tendon has no uniform thickness (McCreesh, Anjum, Crotty and Lewis, 2016), the image was taken between 10 and 15 mm from the LHB for all the subjects (see Figure 2).

Acromiohumeral Distance at 60° of Scaption

The subject was seated in the same initial position as for the AHD at rest. A liquid damped inclinometer was fixed with an elastic band to the arm to define 60° of SPE. Sixty degrees of shoulder elevation has previously been suggested for provocation of subacromial soft tissue (Bailey et al, 2015), since a higher range of elevation or abduction is not considered reliable because of the acoustic shadow of the acromion (Mackenzie, Bdaiwi, Herrington and Cools, 2016; Navarro-Ledesma and Luque-Suarez, 2018). Moreover, patients with impingement syndrome usually present symptoms between 60° and 120° during painful arc and therefore 60° is a clinically relevant angle where the pain usually occurs. The probe and the arm were placed as in the measurement of the AHD at rest, and the examiner assisted the arm elevation of the participant up to 60° of SPE and then the subject had to actively hold the position, in which the US image was taken and measured (see Figure 3).

Acromiohumeral Distance at 60° of Scaption with Weight

The measurement was performed in the same position as for the AHD at 60°, and the subject had to lift a weight of 1 kg if their body weight was inferior to 70 kg, or 2 kg if their body weight was superior to 71 kg (see Figure 4). Rater A indicated to rater B which weight to choose. Since SSP is considered the result of excessive load on RC structures (Lewis, McCreesh, Roy and Ginn, 2015), the aim of this measure was to verify the immediate impact of a free weight on the subacromial space, identified by the AHD on the US image. Although the measurement did not induce a patient-specific fatigue state of the RC, it was related to the body weight.

Occupation Ratio

Occupation ratio was defined as the percentage of subacromial space occupied by the supraspinatus tendon, using the formula: [(STT/AHD)*100] (Michener et al, 2015). It was calculated at 0° (rest), at 60° of SPE, at 60° of SPE with weight.

Data Analysis

Data were analyzed using the SPSS software (Version 26.0). All variables were tested for normality using the Kolmogorov-Smirnov test ($P>0.05$) with Lilliefors Significance Correction. Inter-rater reliability was tested with Intraclass Correlation Coefficient (ICC). The mean of 3 trials for each rater for each US measurement was reported in SPSS and the ICC 2,3 model (2-way random-effects model – absolute agreement, average measure) (Shrout and Fleiss, 1979) with 95% of confidence intervals was used for the calculation of the inter-rater reliability. This model was chosen to generalize the results to any raters with similar characteristics, considering the mean value of 2 raters as an assessment basis for future studies (Koo and Li, 2016). If data were not normally distributed, they were transformed in natural logarithm. However, since differences in the ICC values were inferior than 0.09 between raw and transformed data for both SSP and controls, the ICC of the initial values was used in the analysis. Reliability coefficients were interpreted as follows: excellent if ICC was above 0.90, good if it was between 0.75 and 0.90, moderate if it was between 0.50 and 0.75, poor if it was less than 0.50 (Portney, 2020).

Standard Error of Measurement (SEM) was calculated using the standard deviation (SD) of difference in scores: SEM= SDmean difference/√2 (de Vet, Terwee, Knol and Bouter, 2006). As indicated by de Vet et al, this formula does not count for error due to systematic differences between raters. Bland-Altman plots (Bland and Altman, 1986) with 95% limits of agreement (LOA) were conducted to detect visually systematic bias (i.e. fixed bias) or any possible relationship between the differences and the mean (i.e. proportional bias, funnel effect), considering all subjects. They were further verified as follows: if the mean difference between raters differed significantly from 0 in a one sample t-test, this indicated the presence of systematic bias; linear regression analysis was used to find possible relationship between the difference (as dependent variable) and the mean (as independent variable) for every US measure. Differences between raters in SSP and controls were tested with paired t-test and, if data were not normally distributed, with the Wilcoxon signed-ranks test. The Minimal Detectable Change (MDC) at 95% of confidence interval was finally calculated, using the formula: MDC= SEM×√2×1.96 (Portney, 2020).

RESULTS

Fifty patients were screened, but nine were excluded because they did not meet the eligibility criteria. Therefore, 41 subjects were included: 21 subjects with SSP and 20 asymptomatic controls. Demographics of the two groups (controls and SSP) were comparable (Table 1), except for age and gender, where the SSP group was considerably older ($p<0.001$)

and included significantly more women (p=0.014) compared with the controls. The two groups were different in terms of pain on NPRS and SPADI score ($p<0.0001$), as anticipated, due to the study design. Mean values of AHD at rest and 60° of SPE with or without weights, STT and calculated OR with relative standard deviations, confidence intervals at 95% and ranges for both groups are reported in Table 2.

ICC values are presented in Table 3. ICC for controls was good for AHD at rest and at 60° (0.76-0.77), moderate for STT and AHD at 60° with weights (0.53-0.72), while it was poor for OR (below 0.44). ICC for SSP was moderate for AHD at rest and at 60°, STT, OR at rest and at 60° (0.52-0.74) and poor for AHD at 60° with weights and OR at 60° with weights (0.33-0.36). Inter-rater reliability of AHD was better in controls than SSP in all positions, while it was vice versa for STT. Inter-rater reliability of OR was also higher in the SSP group, except when subjects were holding a free weight (Table 3). The SEM and MDC were always higher in the SSP group than in the control group for all US measures, except for AHD at 60° without weights (Table 3).

From Bland-Altman plots differences between raters in AHD and STT were less than 1 mm, except for AHD at 60°, in which the mean difference was 1.14 mm. Differences in OR were around 5% at rest and they were above 13% when the arm was elevated with or without weights (Figure 5). Considering all subjects, rater A systematically evaluated higher AHD, and lower STT and OR than rater B, which was confirmed in one sample t-test ($p<0.007$). No proportional bias or funnel effect was evident, confirmed by regression analysis that did not find any significant relationship between mean and difference for any US outcome. Ninety-five percent of the measures of AHD at rest and at 60° were within -2.25 and 3.93 mm while the range was wider for AHD at 60° with weights (between -2.60 and 4.48 mm). LOA showed smaller variation for STT, from -1.48 and 0.88 mm. Concerning the OR, LOA ranged between -21 and 11% at rest, while the range enlarged to -51 and 20% at 60° and further more at 60° with weight (-56 and 30%). Differences between raters were found significant for both controls and SSP in all US measures ($p<0.05$), except in OR at 60° with weights in both groups and in AHD at 60° with weights for SSP.

DISCUSSION

The aim of this study was to investigate the inter-rater reliability between two US examiners (experienced examiner and novice examiner) and to explore if inter-rater reliability was different in asymptomatic subjects and subjects with SSP. Considering only asymptomatic subjects, inter-rater reliability was good for AHD at rest and at 60°, moderate for STT and AHD at 60° with weights, while it was poor for OR at rest or 60° with or without weights. Concerning the inter-rater reliability for SSP group, it showed moderate levels for AHD at rest and at 60°, STT, OR at rest and at 60° and it was poor for AHD at 60° with weights and OR at 60° with weights. Higher inter-rater reliability was found in controls for AHD in all positions and for OR at 60° with weights in comparison with SSP group, while it was vice versa for STT and OR at rest or at 60° without weights.

Considering only asymptomatic subjects, the current ICC in AHD showed good inter-rater reliability similarly to previous studies (Desmeules et al, 2004; Mackenzie, Bdaiwi, Herrington and Cools, 2016) but lower than other excellent findings (Hougs Kjaer et al, 2017). Regarding subjects with SSP, AHD at rest was moderately reliable in line with Pijls et al (Pijls et al, 2010) but lower than other studies (Bağcier, Geler Külcü, Yorulmaz and Altunok, 2020; Hougs Kjaer et al, 2017; McCreesh, Anjum, Crotty and Lewis, 2016). When the arm was elevated at 60° in the scapular plane, inter-rater reliability of AHD was good and therefore higher than moderate levels previously reported in asymptomatic subjects (Mackenzie, Bdaiwi, Herrington and Cools, 2016). In contrast, AHD at 60° was similarly moderate when compared to previous measures on symptomatic subjects (Pijls et al, 2010). It is important to note that these preceding studies conducted measures in abduction instead of scapular elevation, which is usually 30-45° from the frontal plane. The current symptomatic subjects showed less reliable results between the two raters compared to asymptomatic subjects for AHD in all positions, with or without weights. The lower reliability in AHD for the SSP group, and especially at 60°, can be attributed partially to the interpretation of the US image, which is more difficult in symptomatic patients because the inferior edge of the acromion is less clearly identifiable as bony landmark (Pijls et al, 2010). This can be as a result of the acoustic shadow of the acromion on the inferior edge or because of the presence of calcifications or fibrotic changes, which makes the visualization of the inferior edge more challenging (Pijls et al, 2010). Moreover, calcifications can also create posterior shadowing (Singh, 2012). In particular in the position of 60° with weight, we obtained moderate inter-rater reliability in asymptomatic subjects and poor inter-rater reliability in symptomatic subjects. It was too painful for some patients to hold the weight for six testing times at 60°. In three patients with SSP, the weights was lowered from 2 to 1 kg or from 1 to 0.5 kg, in order to complete the US measurements. This could have negatively affected the correct positioning and consequently our results. Moreover, it is known that AHD decreases significantly after one hour of fatigue loading exercises in both SSP and asymptomatic subjects (McCreesh, Purtill, Donnelly and Lewis, 2017). Although specific loading exercises were not performed and the US examination lasted less than one hour, we can note that rater B, who always measured after rater A, presented significantly lower values of AHD in all positions in both groups (except for AHD with weights in SSP). Therefore, we can hypothesize that the subjects measured by rater B were more tired because of the maintenance and repeated arm elevations with and without weight. This may have caused superior migration of the humeral head because of the fatigue-state of the RC muscles (Royer et al, 2009), leading to decreased values of AHD for rater B.

Inter-rater reliability was moderate in STT for both asymptomatic or symptomatic subjects but lower than previous findings in both SSP (Bağcier, Geler Külcü, Yorulmaz and Altunok, 2020; Hougs Kjaer et al, 2017; McCreesh, Anjum, Crotty and Lewis, 2016) or control (Hougs Kjaer et al, 2017) groups. Differences in the inter-rater reliability results compared to previous studies might be attributed to different approaches in the US measurement of STT in terms of US views, positioning and landmarks. Firstly, the transverse view is preferred for the calculation of the occupation ratio (Michener et al, 2015), but the longitudinal view is also frequently used in reliability studies (Hougs Kjaer et al, 2017; Ingwersen et al, 2016). Secondly, the patient position varies as well, from maximal internal rotation (Hougs Kjaer et al, 2017) to modified Crass position (Bağcier, Geler Külcü, Yorulmaz and Altunok, 2020; McCreesh, Anjum, Crotty and Lewis, 2016; Michener et al, 2015; Navarro-Ledesma, Struyf, Falla and Luque-Suarez, 2019). Thirdly, the long head of the biceps (LHB) is usually taken as landmark but there is no consensus on how many millimeters should be counted from this landmark to the insertion of supraspinatus tendon to measure STT. Different studies took the mean of 2 or 3 points from 5, 10, 15 and/or 20 mm from the LHB in the transverse view (Bağcier, Geler Külcü, Yorulmaz and Altunok, 2020; McCreesh, Anjum, Crotty and Lewis, 2016; Michener et al, 2015). We reasoned that taking one measurement between 10 and 15 mm was suitable for all patients and that it was more applicable to routine use in clinical practice.

The current ICC of OR at rest was moderate in SSP group, but poor in control group. Being a relative new parameter, only one study evaluated the inter-rater reliability of OR at rest in symptomatic subjects, which was found good and therefore higher than our results (Bağcier, Geler Külcü, Yorulmaz and Altunok, 2020). However, it is important to note that inter-rater reliability of OR depends on the values of AHD and STT, as the OR is a derivate percentage of these values. As a consequence of the higher AHD and lower STT measured by rater A, rater A showed significantly lower OR than rater B at rest or at 60° without weights in both groups. The OR at rest in SSP was around 40% and therefore lower than 50 or 62% found in previous studies (Bağcier, Geler Külcü, Yorulmaz and Altunok, 2020; Michener et al, 2015). However, this was possibly caused by the lower STT measured in the current study.

Bland Altman plots and one sample t-test showed that there was systematic bias as rater A always evaluated AHD higher, and STT and OR lower than rater B. However, the calculated SEM did not include the systematic error and it was, for example in AHD for controls, 0.94 and 1.09 mm at rest and at 60° respectively. This is comparable to SEM of 0.81 mm for AHD at rest and 1.20 mm for AHD at 60° reported by Mackenzie et al, where the angle of 60° was measured in abduction instead of scapular plane elevation (Mackenzie, Bdaiwi, Herrington and Cools, 2016). On the other hand, SEM for AHD at rest was 1.14 mm in symptomatic subjects and therefore higher than other studies, which showed values under 0.5 mm (Bağcier, Geler Külcü, Yorulmaz and Altunok, 2020; McCreesh, Anjum, Crotty and Lewis, 2016). The current MDC was 2.6 mm for AHD at rest in asymptomatic subjects, that is only slightly superior than 2.2 mm reported by Mackenzie et al (Mackenzie, Bdaiwi, Herrington and Cools, 2016). Similarly, the SEM (0.44 mm) and the MDC (1.23 mm) of STT in the symptomatic group were also in line with the results of McCreesh et al (McCreesh, Anjum, Crotty and Lewis, 2016), but higher than the findings of Bagcier et al (Bağcier, Geler Külcü, Yorulmaz and Altunok, 2020). Therefore, our MDC values for AHD at rest or at 60° and for STT in asymptomatic subjects can be used as normative data for future studies. Our MDC for STT in symptomatic population can also be considered for future research, while the MDC for AHD in symptomatic subjects should be interpreted with caution, in light of higher values in comparison with previous studies. However, intra-rater MDC values have been found lower than inter-rater MDC values and a single examiner has been suggested when possible (McCreesh, Anjum, Crotty and Lewis, 2016). Therefore, it is advised to investigate also intra-rater reliability of novice and experienced examiners and relative SEM and MDC values in future studies.

Limitations

Several limitations were present in this study. Firstly, the initial training phase included only three asymptomatic subjects and no symptomatic patients. Moreover, the overall agreement between raters was not tested prior to the start of the inter-reliability phase, as is usually performed in studies with excellent reliability (Hougs Kjaer et al, 2017; Ingwersen et al, 2016). Secondly, it was not possible to compare the different US measures between asymptomatic and SSP patients because of significant differences on confounding factors such as age and gender. Further research is suggested in larger samples with matched controls on these variables. Thirdly, blinding of the US examiners on health status of the participants can be implemented with a third researcher not involved in the measurement who can independently enroll the subjects. Fourthly, it was not always possible to lift the weights for some patients with SSP. Therefore, it is suggested for future studies to start with the lowest weight available (which is usually 0.5 kg) at 60° and then perform the measurement with the first load provoking pain, regardless of the body weight of the person. This would give an indication of load capacity of each individual related to the level of pain and possible relationships between loading at 60° and US measures can be investigated. Lastly, previous studies which involved raters with similar US experience or where a preliminary agreement phase was performed prior to the inter-rater reliability phase, obtained higher inter-rater reliability compared to our results. Therefore, we recommend including an overall agreement phase for future studies, together with a structured training on the specific US measures which have to be analyzed.

CONCLUSION

In conclusion, the inter-rater reliability of this study showed large variability between two US examiners (experienced vs novice) in the measures of AHD and OR at rest or at 60° (with or without weights) and STT, ranging from poor to good.

Clinicians might use the US as additional tool in order to detect subacromial structures and their contribution to pathological mechanisms of SSP. However, the differences between novice examiner and experienced examiner which emerged in this study suggest that US experience plays a relevant role and a structured training including also symptomatic subjects is recommended. Researchers may further investigate the OR in matched case-control studies in different levels of active arm elevation to establish relevant dissimilarities between asymptomatic and symptomatic population. An overall agreement phase before starting the inter-rater reliability is recommended for future studies, in addition to a feasibility phase of the measurements with weights.

ACKNOWLEDGEMENTS

The authors would like to thank PT G.Wollants and PT, PhD K.Kuppens for their help and support in patient recruitment.

FOUNDING DETAILS

This research was supported by the University of Antwerp under the Grant [37063].

DISCLOSURE STATEMENT

The authors report no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author, [filip.struyf@uantwerpen.be], upon reasonable request.

REFERENCES

Bağcier F, Geler Külcü D, Yorulmaz E, Altunok EÇ 2020 Intra- and Inter-Rater Reliability of Ultrasound Measurements of Supraspinatus Tendon Thickness, Acromiohumeral Distance, and Occupation Ratio in Patients With Shoulder Impingement Syndrome. Archives of Rheumatology 35: 385-393.

Bailey LB, Beattie PF, Shanley E, Seitz AL, Thigpen CA 2015 Current rehabilitation applications for shoulder ultrasound imaging. Journal of Orthopaedic and Sports Physical Therapy 45: 394-405.

Bland JM, Altman DG 1986 Statistical Methods for Assessing Agreement between Two Methods of Clinical Measurement. Lancet 1: 307-310.

Bujang MA, Baharum N 2017 A simplified guide to determination of sample size requirements for estimating the value of intraclass correlation coefficient: a review. Archives of Orofacial Sciences 12: 1-11.

de Vet HC, Terwee CB, Knol DL, Bouter LM 2006 When to use agreement versus reliability measures. Journal of Clinical Epidemiology 59: 1033-1039.

Desmeules F, Minville L, Riederer B, Cote CH, Fremont P 2004 Acromio-humeral distance variation measured by ultrasonography and its association with the outcome of rehabilitation for shoulder impingement syndrome. Clinical Journal of Sport Medicine 14: 197-205.

Elvers J, Oostendorp R, Sierevelt I 2003 De Nederlandstalige Shoulder pain and Disability Index (SPADI-Dutch Version) bij patiënten na een subacromiale decompressie volgens Neer: interne consistentie en constructvaliditeit. Nederlands tijdschrift voor fysiotherapie 6: 126-131.

Hougs Kjaer B, Ellegaard K, Wieland I, Warming S, Juul-Kristensen B 2017 Intra-rater and inter-rater reliability of the standardized ultrasound protocol for assessing subacromial structures. Physiotherapy Theory and Practice 33: 398-409.

Ingwersen KG, Hjarbaek J, Eshoej H, Larsen CM, Vobbe J, Juul-Kristensen B 2016 Ultrasound assessment for grading structural tendon changes in supraspinatus tendinopathy: an inter-rater reliability study. BMJ Open 6: e011746.

Koo TK, Li MY 2016 A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. Journal of Chiropractic Medicine 15: 155-163.

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

Kulkarni R, Gibson J, Brownson P, Thomas M, Rangan A, Carr AJ, Rees JL 2015 Subacromial shoulder pain. Shoulder & Elbow 7: 135-143.

Lewis J, McCreesh K, Roy JS, Ginn K 2015 Rotator Cuff Tendinopathy: Navigating the Diagnosis-Management Conundrum. Journal of Orthopaedic and Sports Physical Therapy 45: 923-937.

Mackenzie TA, Bdaiwi AH, Herrington L, Cools A 2016 Inter-rater Reliability of Real-Time Ultrasound to Measure Acromiohumeral Distance. PM & R : the journal of injury, function, and rehabilitation 8: 629-634.

Martinoli C 2010 Musculoskeletal ultrasound: technical guidelines. Insights Imaging 1: 99-141.

McCreesh KM, Anjum S, Crotty JM, Lewis JS 2016 Ultrasound measures of supraspinatus tendon thickness and acromiohumeral distance in rotator cuff tendinopathy are reliable. Journal of Clinical Ultrasound 44: 159-166.

McCreesh KM, Crotty JM, Lewis JS 2015 Acromiohumeral distance measurement in rotator cuff tendinopathy: is there a reliable, clinically applicable method? A systematic review. British Journal of Sports Medicine 49: 298-305.

McCreesh KM, Purtill H, Donnelly AE, Lewis JS 2017 Increased supraspinatus tendon thickness following fatigue loading in rotator cuff tendinopathy: potential implications for exercise therapy. BMJ open sport & exercise medicine 3: e000279.

Michener LA, Subasi Yesilyaprak SS, Seitz AL, Timmons MK, Walsworth MK 2015 Supraspinatus tendon and subacromial space parameters measured on ultrasonographic imaging in subacromial impingement syndrome. Knee Surgery, Sports Traumatology, Arthroscopy 23: 363-369.

Michener LA, Walsworth MK, Doukas WC, Murphy KP 2009 Reliability and diagnostic accuracy of 5 physical examination tests and combination of tests for subacromial impingement. Archives of Physical Medicine and Rehabilitation 90: 1898-1903.

Mintken PE, Glynn P, Cleland JA 2009 Psychometric properties of the shortened disabilities of the Arm, Shoulder, and Hand Questionnaire (QuickDASH) and Numeric Pain Rating Scale in patients with shoulder pain. Journal of Shoulder and Elbow Surgery 18: 920-926.

Navarro-Ledesma S, Luque-Suarez A 2018 Comparison of acromiohumeral distance in symptomatic and asymptomatic patient shoulders and those of healthy controls. Clinical Biomechanics (Bristol, Avon) 53: 101-106.

Navarro-Ledesma S, Struyf F, Falla D, Luque-Suarez A 2019 Non-traumatic chronic shoulder pain is not associated with changes in rotator cuff interval tendon thickness. Clinical Biomechanics (Bristol, Avon) 63: 147-152.

Neer CS, 2nd 1972 Anterior acromioplasty for the chronic impingement syndrome in the shoulder: a preliminary report. Journal of Bone and Joint Surgery (American Volume) 54: 41-50.

Pijls BG, Kok FP, Penning LI, Guldemond NA, Arens HJ 2010 Reliability study of the sonographic measurement of the acromiohumeral distance in symptomatic patients. Journal of Clinical Ultrasound 38: 128-134.

Portney LG 2020 Chapter 32: Measurement Revisited: Reliability and Validity Statistics. In: F.A. Davis (Ed) Foundations of Clinical Research: Applications to Evidence-Based Practice pp. 485-508. F. A. Davis Company - Philadelphia.

Roach KE, Budiman-Mak E, Songsiridej N, Lertratanakul Y 1991 Development of a shoulder pain and disability index. Arthritis Care and Research 4: 143-149.

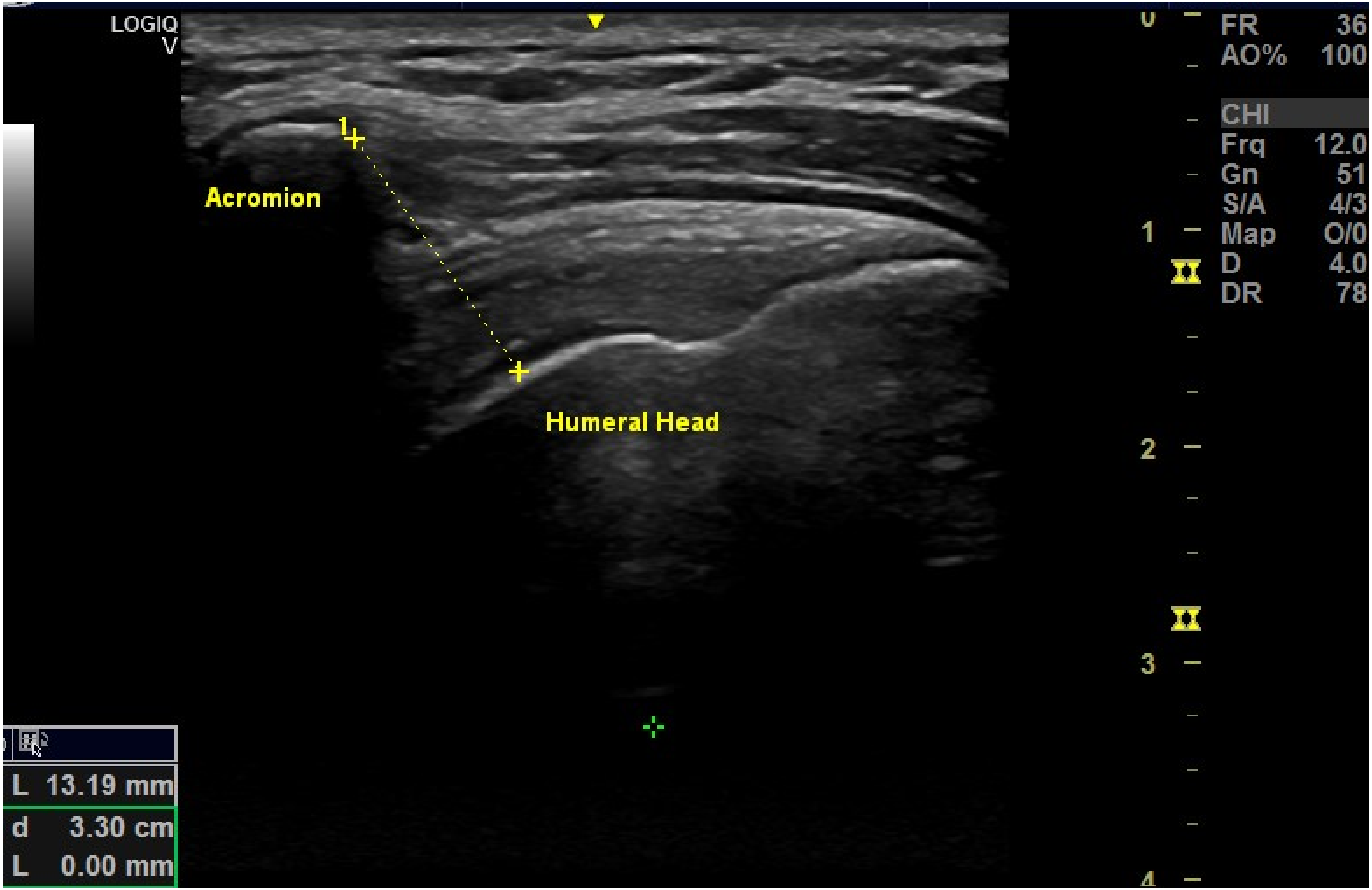
Royer PJ, Kane EJ, Parks KE, Morrow JC, Moravec RR, Christie DS, Teyhen DS 2009 Fluoroscopic assessment of rotator cuff fatigue on glenohumeral arthrokinematics in shoulder impingement syndrome. Journal of Shoulder and Elbow Surgery 18: 968-975.

Savoie A, Mercier C, Desmeules F, Fremont P, Roy JS 2015 Effects of a movement training oriented rehabilitation program on symptoms, functional limitations and acromiohumeral distance in individuals with subacromial pain syndrome. Manual Therapy 20: 703-708.

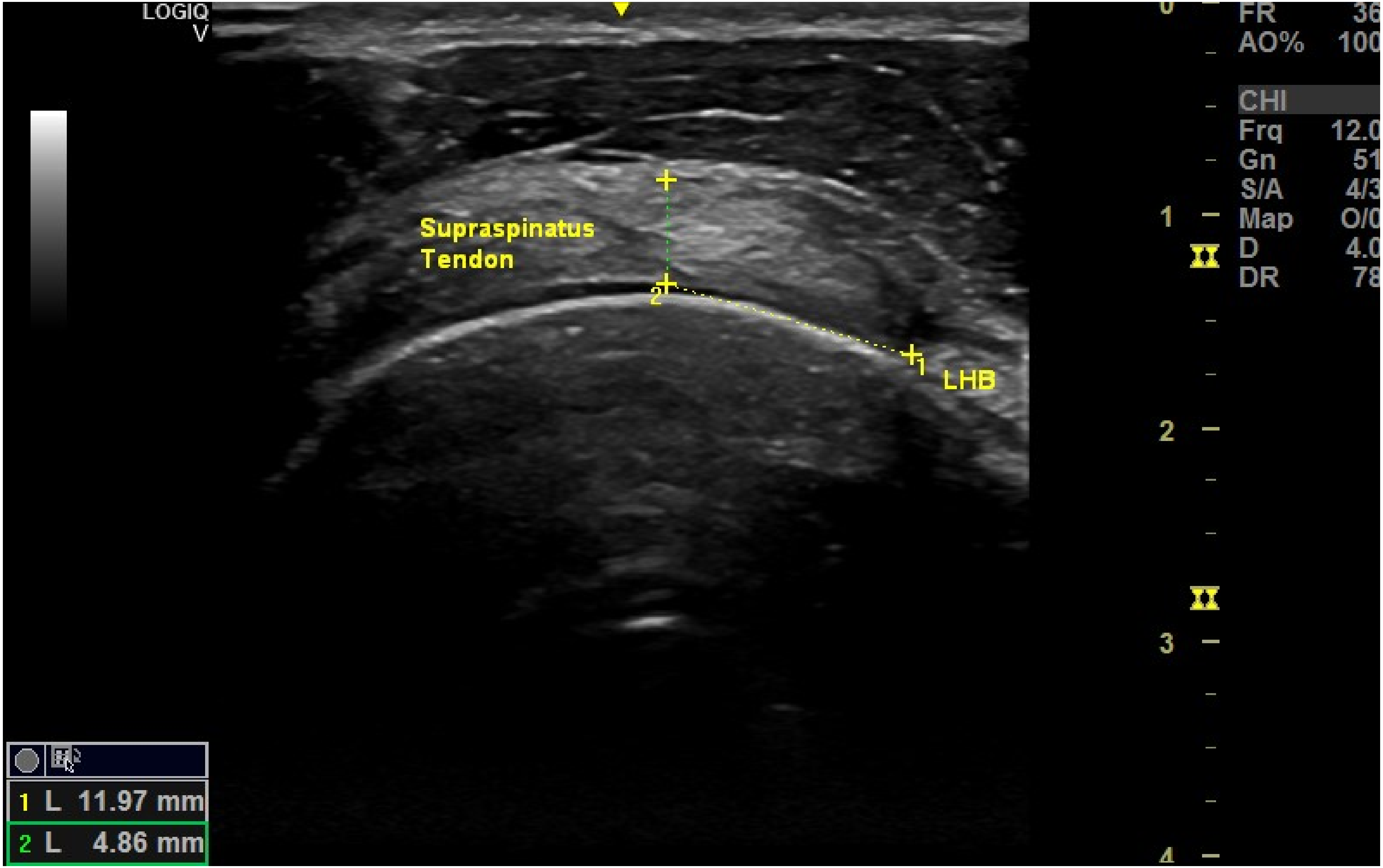
Shrout PE, Fleiss JL 1979 Intraclass correlations: uses in assessing rater reliability. Psychological Bulletin 86: 420-428.
Singh JP 2012 Shoulder ultrasound: What you need to know. Indian Journal of Radiology & Imaging 22: 284-292.
Thoomes-de Graaf M, Scholten-Peeters GG, Duijn E, Karel Y, Koes BW, Verhagen AP 2015 The Dutch Shoulder Pain and Disability Index (SPADI): a reliability and validation study. Quality of Life Research 24: 1515-1519.
Virta L, Joranger P, Brox JI, Eriksson R 2012 Costs of shoulder pain and resource use in primary health care: a cost-of-illness study in Sweden. BMC Musculoskeletal Disorders 13.

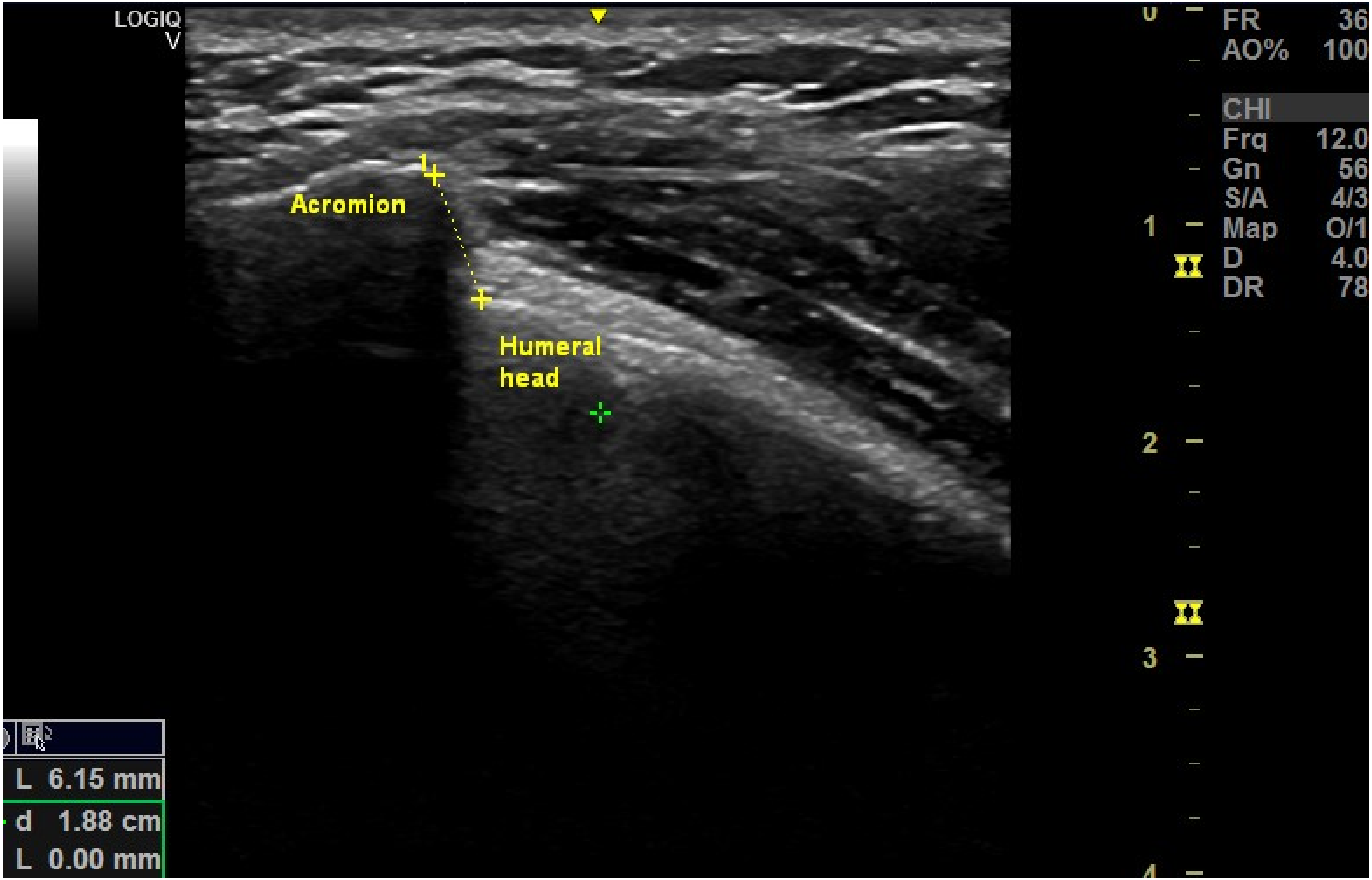
TABLES

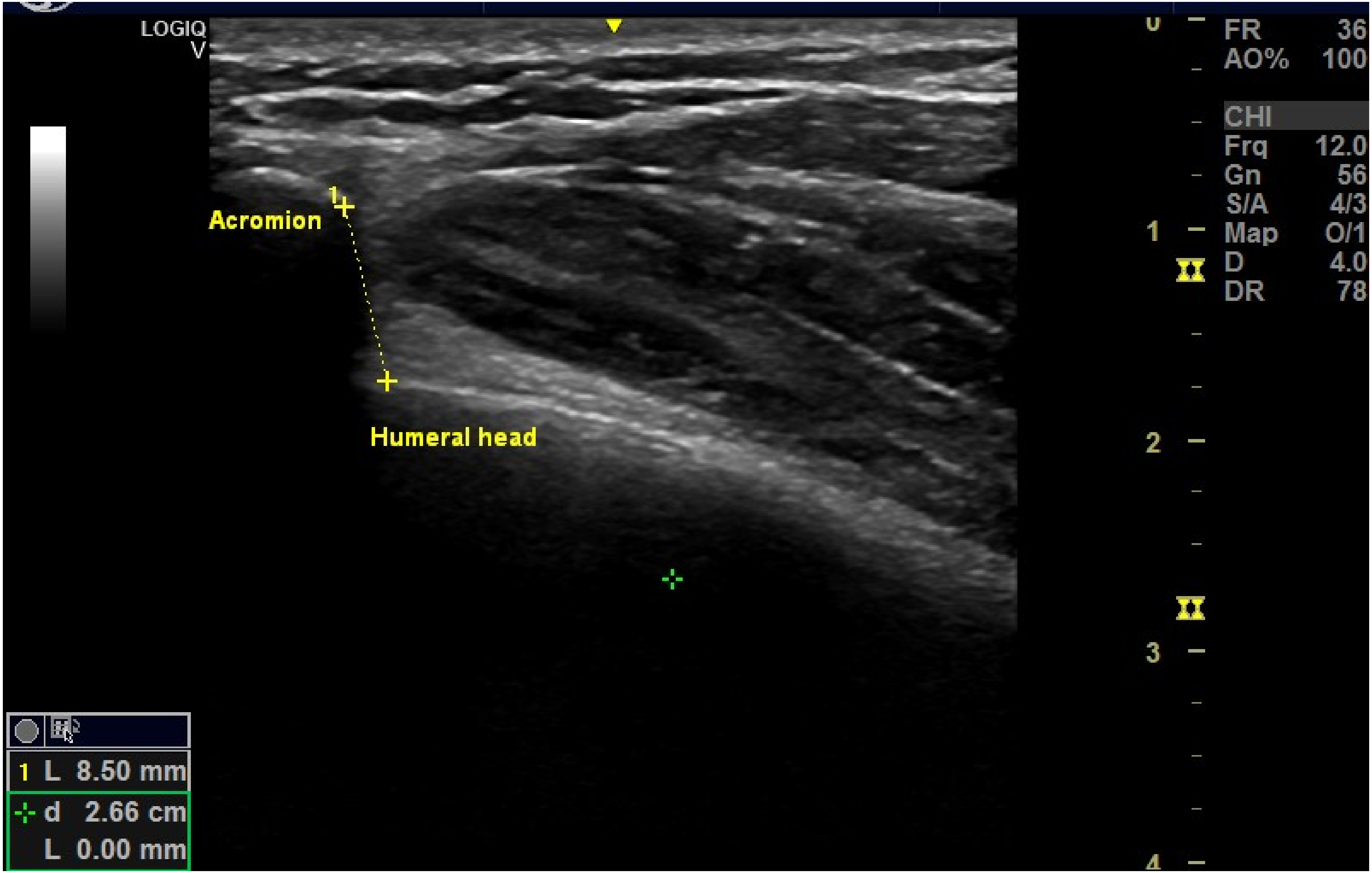
FIGURES



LOGIQ
V
FR 36
AO% 100
CHI
Frq 12.0
Gn 51
S/A 4/3
Map O/0
D 4.0
DR 78
Acromion
Humeral Head
1
2
3
4
L 13.19 mm
d 3.30 cm
L 0.00 mm


LOGIQ
V
Supraspinatus
Tendon
LHB
FR 36
AO% 100
CHI
Frq 12.0
Gn 51
S/A 4/3
Map O/0
D 4.0
DR 78
1 L 11.97 mm
2 L 4.86 mm


LOGIQ
V
Acromion
Humeral
head
FR 36
AO% 100
CHI
Frq 12.0
Gn 56
S/A 4/3
Map O/1
D 4.0
DR 78
1
2
3
L 6.15 mm
d 1.88 cm
L 0.00 mm



LOGIQ
V
Acromion
1
Humeral head
FR 36
AO% 100
CHI
Frq 12.0
Gn 56
S/A 4/3
Map O/1
D 4.0
DR 78
0
1
2
3
4
1 L 8.50 mm
d 2.66 cm
L 0.00 mm

Figure 5. Bland Altman plots for inter-rater reliability of the total group in AHD at rest (Fig 5.1), AHD at 60° (Fig 5.2), AHD at 60° with weight (Fig 5.3), STT (Fig 5.4), OR at rest (Fig 5.5), OR at 60° (Fig 5.6), OR at 60° with weight (Fig 5.7).

Figure 5.1 AHD at rest (in mm)

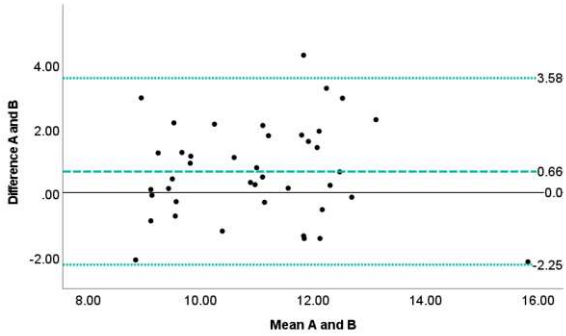


Figure 5.2 AHD at 60° (in mm)

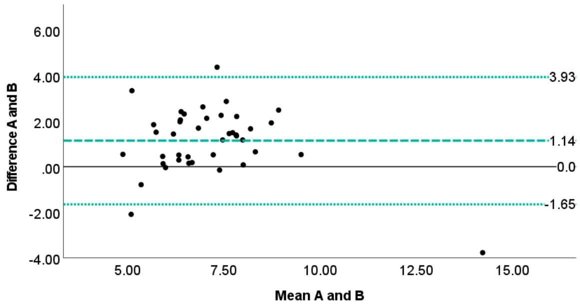


Figure 5.3 AHD at 60° with weight (in mm)

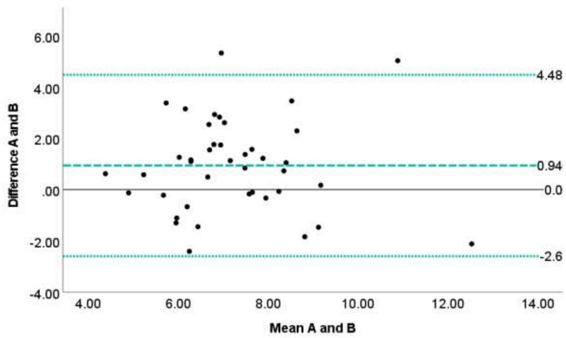


Figure 5.4 STT (in mm)

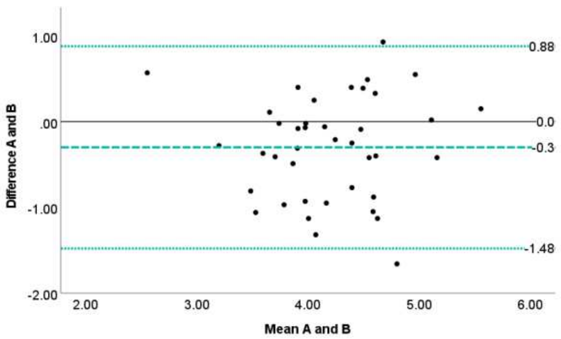


Figure 5.5 OR at rest (in %)

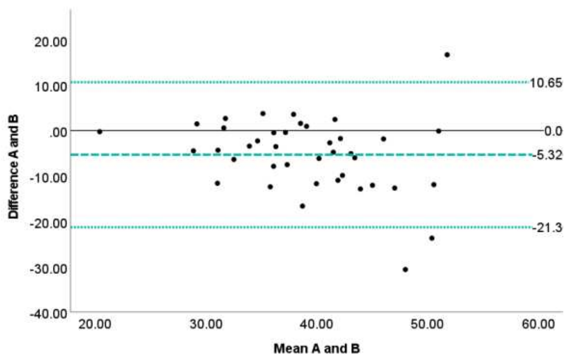


Figure 5.6 OR at 60° (in %)

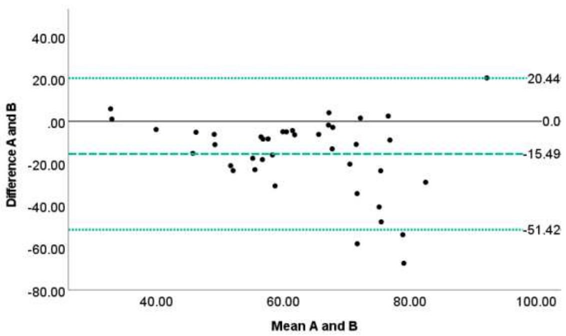
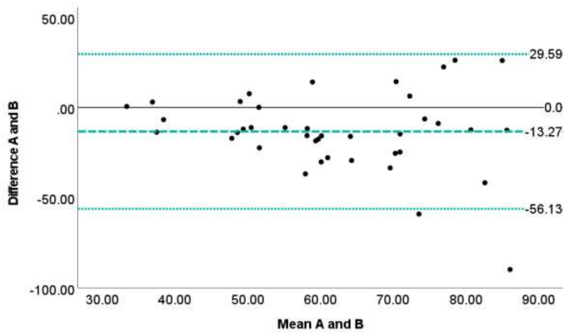


Figure 5.7 OR at 60° with weight (in %)



Legend. *The difference for each US outcome is displayed as: rater A-rater B. Dashed green line= mean difference, dotted green line= limits of agreement (LOA), solid line= perfect agreement (no difference between raters).* ***Abbreviations***. *AHD=AcromioHumeral Distance, OR=Occupation Ratio, STT=Supraspinatus Tendon Thickness*

TABLES

Table 1. Demographics for the control group and the group with SSP

	Controls (n=20)	SSP (n=21)	p-value
Age in years - mean (SD)	28.00 (3.68)	50.24 (10.69)	<0.0001
Gender - male/female	10/10	3/18	=0.014
BMI in kg/m^2 - mean (SD)	22.77 (2.69)	22.42 (3.07)	=0.419
Overhead sports - yes/no	8/12	7/14	=0.658
Pain at rest in NPRS - mean (SD)	0.00 (0.00)	1.81 (1.83)	<0.0001
Pain during activity in NPRS – mean (SD)	0.00 (0.00)	4.29 (2.51)	<0.0001
Pain during past week in NPRS – mean (SD)	0.05 (0.22)	5.14 (2.22)	<0.0001
SPADI score – mean (SD)	0.1 (0.45)	42.93 (22.37)	<0.0001

***Abbreviations**: BMI= Body Mass Index, NPRS= Numeric Pain Rating Scale, SD= Standard Deviation, SPADI= Shoulder Pain and Disability Index, SSP= Subacromial Shoulder Pain*

Table 2. Descriptive statistics of the seven US measures for the control group and the group with SSP

	Rater A			**Rater B**		
	Mean (SD)	95%CI	Range	Mean (SD)	95%CI	Range
CONTROLS (n=20)						
AHD0 (mm)	11.76 (1.36)	11.12-12.40	9.72-14.74	10.81 (2.06)	9.85-11.78	7.47-16.90
STT (mm)*	3.95 (0.61)	3.65-4.24	3.00-5.14	4.25 (0.48)	4.03-4.48	3.60-5.11
AHD60 (mm)	7.98 (1.55)	7.26-8.71	5.97-12.35	6.70 (2.49)	5.53-7.86	3.45-16.12
AHD60w (mm)	7.96 (1.87)	7.09-8.84	4.84-13.40	6.73 (2.17)	5.71-7.75	4.05-13.59
OR0 (%)*	33.94 (5.42)	31.32-36.55	25.23-45.11	40.67 (8.59)	36.53-44.81	28.46-63.32
OR60 (%)*	50.85 (7.90)	47.04-54.66	35.71-66.28	69.47 (19.14)	60.24-78.70	29.84-112.75
OR60w (%)*	51.36 (10.10)	46.49-56.23	30.67-71.79	69.31 (18.15)	60.56-78.10	35.39-103.46
SSP (n=21)						
AHD0 (mm)	10.90 (1.82)	10.07-11.72	7.80-14.26	10.50 (1.14)	9.99-11.02	9.04-12.56
STT (mm)	4.14 (0.70)	3.83-4.46	2.84-5.63	4.43 (0.75)	4.09-4.77	2.27-5.63
AHD60 (mm)	7.42 (1.53)	6.73-8.11	4.05-9.78	6.42 (1.08)	5.93-6.92	4.61-9.25
AHD60w (mm)	7.46 (1.67)	6.70-8.22	4.70-10.26	6.80 (1.39)	6.17-7.44	4.08-9.74
OR0 (%)	38.73 (7.60)	35.28-42.19	20.29-60.13	42.78 (9.16)	38.61-46.95	20.54-62.26
OR60 (%)	58.55 (16.60)	50.99-66.10	33.49-102.47	71.20 (18.05)	62.99-79.42	32.47-105.83
OR60w (%)	59.28 (19.99)	50.18-68.38	33.69-98.02	68.32 (20.70)	58.90-77.74	33.09-130.93

Abbreviations*: 95% CI= 95% Confidence Interval, AHD0= AcromioHumeral Distance at 0 degrees, AHD60= AcromioHumeral Distance at 60 degrees, AHD60w= AcromioHumeral Distance at 60 degrees with weight, OR0= Occupation Ratio at 0 degrees, OR60= Occupation Ratio at 60 degrees, OR60w= Occupation Ratio at 60 degrees with weight, STT= Supraspinatus Tendon Thickness, SD= Standard Deviation, SSP= Subacromial Shoulder Pain. *One subject is missing.*

Table 3. Inter-rater reliability of seven US measures for the control group and the group with SSP

	ICC (95% CI)	Mean difference (SD)	SEM	MDC_{95}	$MDC_{\%}$
CONTROLS (n=20)					
AHD0	0.77 (0.29;0.92)	0.95 (1.33)	0.94	2.60	23.04
STT*	0.53 (-0.10;0.81)	-0.31 (0.59)	0.42	1.16	28.38
AHD60	0.76 (0.18;0.92)	1.29 (1.54)	1.09	3.02	41.08
AHD60w	0.72 (0.18;0.89)	1.23 (1.70)	1.20	3.33	45.35
OR0*	0.44 (-0.21;0.77)	-6.74 (7.85)	5.55	15.39	41.26
OR60*	0.23 (-0.30;0.63)	-18.62 (18.19)	12.86	35.65	59.26
OR60w*	0.37 (-0.25;0.73)	-17.95 (16.32)	11.54	32.00	53.03
SSP (n=21)					
AHD0	0.61 (0.05;0.84)	0.39 (1.61)	1.14	3.15	28.93
STT	0.74 (0.37;0.89)	-0.29 (0.63)	0.44	1.23	29.67
AHD60	0.57 (-0.05;0.82)	1.00 (1.33)	0.94	2.60	35.01
AHD60w	0.36 (-0.46;0.73)	0.66 (1.90)	1.35	3.73	50.01
OR0	0.63 (0.14;0.85)	-4.04 (8.39)	5.93	16.45	40.35
OR60	0.52 (-0.09;0.80)	-12.66 (18.44)	13.04	36.14	55.70
OR60w	0.33 (-0.52;0.72)	-9.03 (25.55)	18.07	50.08	78.50

The mean difference, SEM and MDC_{95} are reported in mm for AHD0, STT, AHD60, AHD60w, and they are reported in % for OR0, OR60 and OR60w. ***Abbreviations:*** *AHD0=AcromioHumeral Distance at 0 degrees, AHD60=AcromioHumeral Distance at 60 degrees, AHD60w= AcromioHumeral Distance at 60 degrees with weight, ICC (95% CI)= Intraclass Correlation Coefficient with 95% of CI, MDC_{95}= Minimal Detectable Change at 95% of CI ~~(in mm)~~, $MDC_{\%}$= Minimal Detectable Change in % respect to mean value of raters A and B, OR0= Occupation Ratio at 0 degrees, OR60= Occupation Ratio at 60 degrees, OR60w= Occupation Ratio at 60 degrees with weight, STT= Supraspinatus Tendon Thickness, SEM= Standard Error of Measurement, SD=Standard Deviation, SSP=Subacromial Shoulder Pain. *One subject is missing.*