SHORT COMMUNICATION

RELIABILITY OF ULTRASOUND EVALUATION OF THE LONG HEAD OF THE BICEPS TENDON

Pascale Drolet, MD[1,2], Anne Martineau, MD[1,2], Rémi Lacroix, MD[1,2] and
Jean-Sébastien Roy, PhD[3,4]

*From the [1]Institut de Réadaptation en Déficience Physique de Québec (IRDPQ), [2]CHU de Québec,
Hôpital de l'Enfant-Jésus, [3]Centre Interdisciplinaire de Recherche en Réadaptation et Intégration Sociale
(CIRRIS) and [4]Département de Réadaptation, Faculté de Médecine, Université Laval, Québec, Canada*

Objective: To determine the reliability of quantitative measures of the long head of the biceps tendon using an ultrasound-imaging system.
Design: Intra- and inter-rater reliability study.
Subjects/patients: Thirty-one participants without shoulder pain.
Methods: All participants took part in 3 ultrasound imaging sessions; they were assessed by 2 evaluators (inter-rater reliability), one of whom assessed them twice (intra-rater reliability). All measurements were taken at the widest identified part of the tendon using longitudinal and transverse views. Measurements of the long head of the biceps tendon included width, thickness and cross-sectional area. Intra-class correlation coefficients and minimal detectable change were used to characterize reliability.
Results: Intra- and inter-rater reliabilities were excellent for all measures when the mean of 2 measures were considered, except for inter-rater reliability of the width, for which it ranged from 0.76 to 0.86. Minimal detectable change ranged from 0.3 to 1.6 mm for width and thickness, and from 2.8 to 4.9 mm² for cross-sectional area.
Conclusion: Ultrasound measurement of the long head of the biceps tendon is a highly reliable method, except for the width. When measuring the long head of the biceps tendon, a mean of 2 measurements is recommended. Now that reliability has been shown in healthy individuals, the next step will be to determine the validity/reliability of these quantitative measures in symptomatic shoulders.

INTRODUCTION

The long head of the biceps tendon (LHBT) is a common source of shoulder pain (1, 2). A close relationship has been shown between rotator cuff tears (which occur in up to 50% of the population) and associated LHBT injuries (3, 4). In fact, microscopic chronic inflammation and gross degeneration of the LHBT has been observed in more than 70% of shoulders with either partial- or full-thickness rotator cuff tears (4). Although relatively few studies have evaluated the diagnostic accuracy of ultrasound for the evaluation of the LHBT, results demonstrate the relevance of ultrasound evaluation in clinical settings, as ultrasound is accurate for evaluating dislocation or subluxation of the LHBT (5–7), as well as LHBT full-thickness tears (8). On the other hand, studies have shown that ultrasound is not sensitive (Sn = 0.27) when diagnosing partial rupture, tendinosis and tenosynovitis (8). The diagnosis of partial rupture, tendinosis and tenosynovitis of the LHBT is based on a qualitative analysis of the tendon, which implies subjectivity by the evaluator compared with quantitative measures (width, thickness, cross-sectional area (CSA)). These quantitative measures could be used for diagnosis of LHBT disorders instead of qualitative analysis of the tendon if they were shown to be reliable and if they increase diagnostic accuracy. One study has evaluated the reliability of quantitative ultrasound measurements of the LHBT in 20 subjects and found good inter-rater reliability coefficients for LHBT thickness (9). However, the evaluators used a steel reference marker that is not used in clinics. Good to excellent reliability indices have also been shown for quantitative ultrasound measurements of patellar tendon thickness (10).

Ultrasound is increasingly used for the evaluation of musculoskeletal injuries because it has several advantages: it is less costly than magnetic resonance imaging, does not have the ionizing radiation effects of radiography, and allows the evaluation of the soft tissue structures. Given the high prevalence of LHBT injuries, we believe that it is important to have a better understanding of the psychometric properties of quantitative ultrasound measures of the LHBT. The main objectives of this study are to determine the intra- and inter-rater reliability of ultrasound measures of width, thickness and CSA of the LHBT in people with uninjured shoulders, as well as the impact of the number of measurements on the reliability level.

METHODS

Participants were recruited through advertisements in a research centre (convenience sample). Inclusion criteria were: age between 18 and 70

years old; no shoulder pain or limitations; and no pain or weakness in the following tests: (*i*) Neer, Hawkins-Kennedy, Jobe, Yergason, Speed; and (*ii*) resisted shoulder external rotation or abduction, or elbow supination or flexion. Exclusion criteria were: fractures, shoulder surgery, cervicobrachialgia. This study was approved by the ethics committee of the Quebec Rehabilitation Institute.

Measurement protocol

Participant characteristics (sex, age, dominance, height, weight) were collected. Thereafter, ultrasonographic measurements of the LHBT were conducted using a MyLab®Five (Biosound Esaote, Italy) ultrasound scanner with a 7.5–12 MHz linear array probe. Ultrasound parameters, such as image field depth (5 cm), gain (58%) and frequency (12 MHz) were established during pilot testing and were identical for all participants. Participants took part in 3 ultrasound-imaging sessions on the same day; they were assessed on both shoulders by 2 evaluators (inter-rater reliability), 1 of whom performed the assessment twice (intra-rater reliability). The 2 evaluators were physical medicine and rehabilitation residents with minimal experience in ultrasound imaging.

Ultrasound images were collected in a sitting position, with the arm at rest on the lap and the elbow flexed, forearm supinated and wrist in a neutral position. The linear probe was placed on the anterior face of the shoulders (on the LHBT), perpendicular to the humerus. Evaluators proceeded to scan the LHBT transversally from the pectoralis major tendon inferiorly to the rotator interval superiorly (6). All measurements were taken at the widest visually and quantitatively identifiable part of the LHBT. Measurements were taken with a transverse view, and then a longitudinal view, which was obtained by rotating the probe 90° while remaining at the same level as the transverse view. This sequence was repeated 3 times by the same evaluator, since the first measurement, the mean of the 2 first measurements and the mean of the 3 measurements were used for reliability analyses. The same

protocol was then applied to the other shoulder, which completed the first imaging session (E1). The same sequence was then performed a second time (E2) by the second evaluator. After a 5-min rest period, a third and final imaging session (E3) was performed by one of the evaluators (one performed 16 re-evaluations at E3, the other 15 re-evaluations). Using transversal images, width, thickness and CSA were measured (Fig. 1a). Thickness was also measured using the longitudinal images (Fig. 1b). Measurements of width, thickness and CSA (using a computer-assisted tool for CSA) were made by the evaluators after the 3 ultrasound imaging sessions, but compiled by another person. The evaluators were blinded to the images and measurements obtained by the other evaluator and to all of their previous measurements.

Statistical analysis

Participants' characteristics were summarized using descriptive statistics. Means and standard deviations (SD) were calculated for the first measurement, the first 2 measurements and the 3 measurements of the 3 ultrasound-imaging sessions for each LHBT measure. Intra-rater reliability was analysed by comparing the first measurements, the mean of the first 2 measurements and the mean of the 3 measurements of each US images of E1 with those of E3. For inter-rater reliability, the first measurement of the first rater was compared with the second rater's first measurement (E1 and E2). The same intersession comparison was made using the mean of the first 2 measurements and the mean of the 3 measurements. Reliability indices were calculated separately for both shoulders.

White tests were used to estimate homoscedasticity of data before each analysis. Intraclass Correlation Coefficient (ICC) (2-way mixed model) and 95% confidence interval (95% CI) were calculated to assess intra- and inter-rater reliability (11). ICC values were considered excellent > 0.90, good from 0.75 to 0.90, fair from 0.40 to 0.74, or poor < 0.40 (12). Absolute reliability was assessed with minimal detectable

Fig. 1. Position of the probe, ultrasound image and quantitative measures of the long head of the biceps in the (A) transverse view and (B) longitudinal view. With the transverse views, the width (AB distance), thickness (CD distance) and cross-sectional area (CSA) were measured. With the longitudinal views, the thickness (CD distance) was measured.
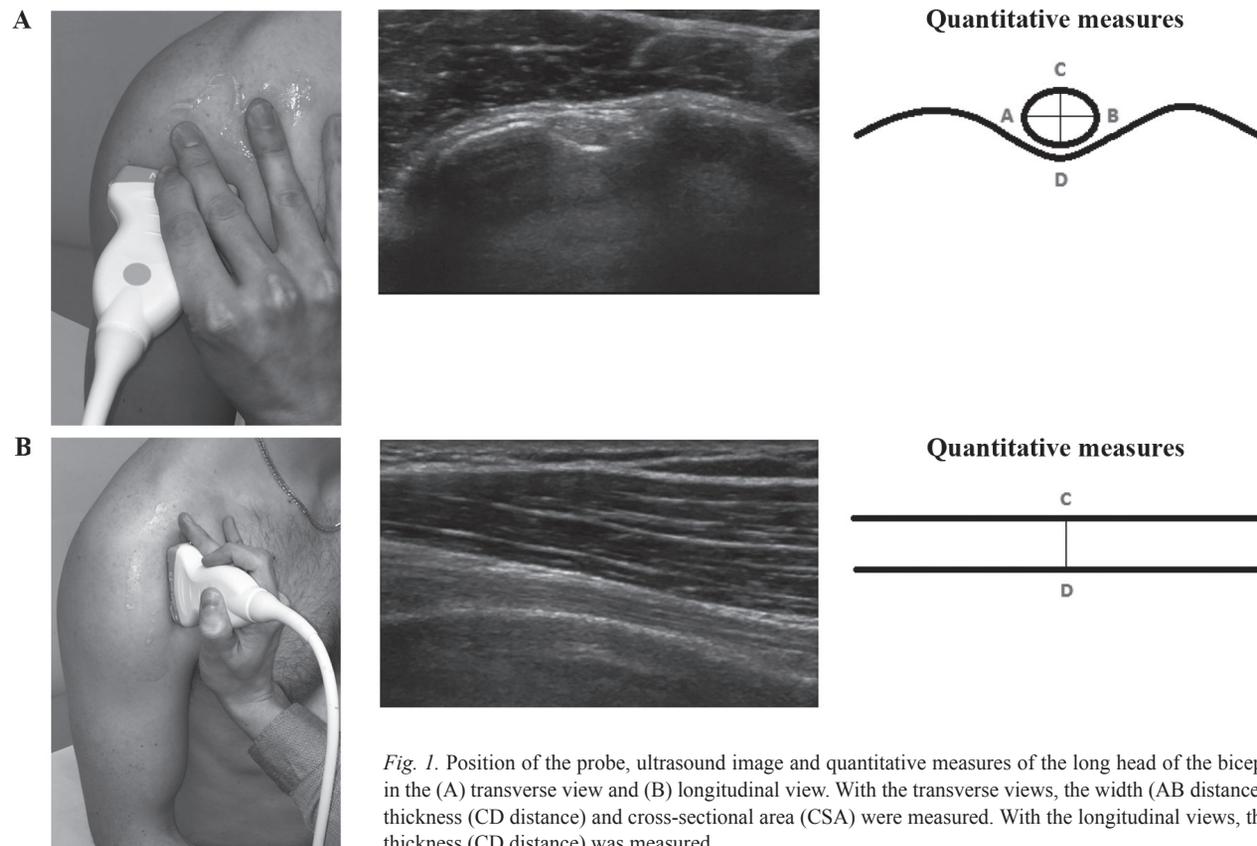
Table I. *Quantitative measures (mean of 2 measurements) of the long head of the biceps*

		Evaluator 1 (*n*=31)		Evaluator 2 (*n*=31)	
		Left shoulder Mean (SD)	Right shoulder Mean (SD)	Left shoulder Mean (SD)	Right shoulder Mean (SD)
Measure					
Transverse view	Width (mm)	4.8 (1.2)	4.8 (1.0)	4.7 (0.6)	4.8 (0.8)
	Thickness (mm)	2.6 (1.4)*	2.5 (1.3)*	2.8 (1.3)*	2.8 (1.3)*
	CSA (mm²)	12.7 (7.7)	12.3 (8.0)	12.1 (6.6)	12.3 (7.4)
Longitudinal view	Thickness (mm)	4.1 (1.1)*	4.1 (1.1)*	4.1 (1.1)*	3.9 (1.2)*

*Significant difference ($p<0.05$) between thickness in the transverse and longitudinal views.
CSA: cross-sectional area; SD: standard deviation.

change (MDC) at 95% CI (11). The MDC can be used to determine whether the change is statistically meaningful (i.e. to determine the measurement error). Agreement within and between raters was determined using the Bland-Altman plotting method (13). Eventual systematic biases between measures of thickness using longitudinal and transverse views were tested with paired *t*-tests and linear regressions. A significant mean difference between thickness in longitudinal and transverse views would indicate a bias, while a regression slope statistically different from 1 would reveal that the bias may not be systematic. Analyses were completed with Statistical Package for the Social Sciences (SPSS) version 23.0 (IBM Corp., Armonk, NY, USA).

RESULTS

Thirty-one participants were included in the study (19 women, 12 men; 29 right-handed, 2 left-handed; mean age 39.0 years (standard deviation (SD) 16.4); mean height 168.2 cm (SD 11.4); mean weight 66.5 kg (SD 9.9)). The means of the 4 US measures are shown in Table I. No deviation of heteroscedasticity was detected for all reliability analyses; therefore, the variances were considered homogeneous. For intra-rater analyses, the reliability was good to excellent (0.77–0.96) when the first measurement was considered, while it was excellent (0.90–0.99) when the mean

of the first 2 or 3 measurements were assessed (Table II). For inter-rater analyses, reliability was excellent, with ICC superior to 0.9, except for the width, for which it ranged from fair to good (0.64–0.89) (Table III). For intra- and inter-rater measurements, the MDC ranged from 0.3 to 1.6 mm for width and thickness, and from 2.8 to 7.2 mm² for CSA. Bland-Altman plots revealed that differences for intra- and inter-rater analyses of all US measures were centred around zero (i.e. no bias indicated).

Fig. 1b shows that the 2 measures of thickness (longitudinal and transverse views) were not centred on the dotted line (the dotted line represents an absolute association between both measures) and that the 95% CI for all 4 slopes (2 evaluators * 2 shoulders) included the value 1. Furthermore, a significant difference (mean difference between 1.3 and 1.7 mm) was observed between the mean thickness in the longitudinal and transverse views (Table I).

DISCUSSION

This study was designed to investigate the reliability of quantitative measures of the LHBT. Intra- and inter-rater reliability

Table II. *Intra-rater reliability of the quantitative measures of the long head of the biceps*

	Evaluator 1 and 2 combined (*n*=31)			
	Left shoulder		Right shoulder	
Measures	ICC (95% CI)	MDC	ICC (95% CI)	MDC
Transverse view				
Width, mm				
First measurement	0.88 (0.76–0.94)	1.0	0.77 (0.58–0.88)	1.4
Mean of first 2 measurements	0.93 (0.85–0.96)	0.9	0.95 (0.89–0.97)	0.9
Mean of the 3 measurements	0.90 (0.78–0.95)	1.1	0.96 (0.92–0.98)	0.8
Thickness, mm				
First measurement	0.93 (0.86–0.97)	0.9	0.94 (0.89–0.97)	0.8
Mean of first 2 measurements	0.99 (0.98–1.00)	0.5	0.99 (0.98–1.00)	0.5
Mean of the 3 measurements	0.99 (0.98–1.00)	0.5	0.99 (0.98–1.00)	0.5
CSA, mm²				
First measurement	0.95 (0.90–0.98)	4.1	0.95 (0.90–0.96)	4.6
Mean of first 2 measurements	0.98 (0.96–0.99)	3.5	0.98 (0.97–0.99)	3.6
Mean of the measurements	0.98 (0.95–0.99)	3.8	0.98 (0.97–0.99)	3.7
Longitudinal view				
Thickness, mm				
First measurement	0.92 (0.83–0.96)	0.9	0.96 (0.92–0.98)	0.7
Mean of first 2 measurements	0.98 (0.95–0.99)	0.6	0.99 (0.98–0.99)	0.5
Mean of the 3 measurements	0.94 (0.88–0.97)	1.1	0.99 (0.98–1.00)	0.5

CSA: cross-sectional area; ICC: intraclass correlation coefficients; MDC: minimal detectable change; CI: confidence interval.

Table III. *Inter-rater reliability of the quantitative measures of the long head of the biceps*

Measures	Left shoulder (*n*=31)		Right shoulder (*n*=31)	
	ICC (95% CI)	MDC	ICC (95% CI)	MDC
Transverse view				
Width, mm				
First measurement	0.65 (0.38–0.81)	1.6	0.64 (0.37–0.81)	1.7
Mean of first 2 measurements	0.76 (0.50–0.88)	1.6	0.86 (0.72–0.93)	1.2
Mean of the 3 measurements	0.85 (0.69–0.93)	1.3	0.89 (0.77–0.95)	1.1
Thickness, mm				
First measurement	0.92 (0.85–0.96)	1.0	0.93 (0.83–0.97)	1.0
Mean of first 2 measurements	0.98 (0.96–0.99)	0.7	0.99 (0.98–1.00)	0.5
Mean of the 3 measurements	0.98 (0.97–0.99)	0.7	0.98 (0.97–0.99)	0.6
CSA, mm^2				
First measurement	0.91 (0.83–0.96)	5.6	0.90 (0.81–0.95)	7.2
Mean of first 2 measurements	0.97 (0.94–0.99)	4.9	0.99 (0.98–0.99)	3.3
Mean of the 3 measurements	0.98 (0.96–0.99)	3.8	0.99 (0.98–0.99)	3.2
Longitudinal view				
Thickness, mm				
First measurement	0.93 (0.86–0.97)	0.6	0.91 (0.84–0.95)	0.9
Mean of first 2 measurements	0.98 (0.96–0.99)	0.6	0.96 (0.92–0.98)	0.9
Mean of the 3 measurements	0.94 (0.88–0.97)	1.3	0.97 (0.95–0.99)	0.7

CSA: cross-sectional area; ICC: intraclass correlation coefficients; MDC: minimal detectable change; CI: confidence interval.

was excellent for all measurements except for width, which can be classified as fair to excellent. These results are similar to those of Collinger et al. (9), who found that inter-rater reliability was good for LHBT thickness with a coefficient of dependability (Φ) superior to 0.80. Skou & Aalkjaer (10) had similar findings for intra- and inter-rater reliability for patellar tendon thickness (ICC >0.70). They found excellent reliability (intra- and inter-rater) when using a mean of 2 or 3 measurements (though adding a third measurement did not improve reliability), while lower reliability was obtained when

using 1 measurement. Like Skou & Aalkjaer (10), our study did not find any improvement in reliability when adding a third measurement, and found lower reliability when using 1 measurement (especially for width). MDCs were also consistently higher when 1 measurement was used, while they were similar for the mean of 2 or 3 measurements. Therefore, we recommend using the mean of 2 measurements in daily practice. The fact that medical residents with minimal experience in ultrasound-imaging captured the images was initially seen as a limitation of this study. However, as shown by Wallwork
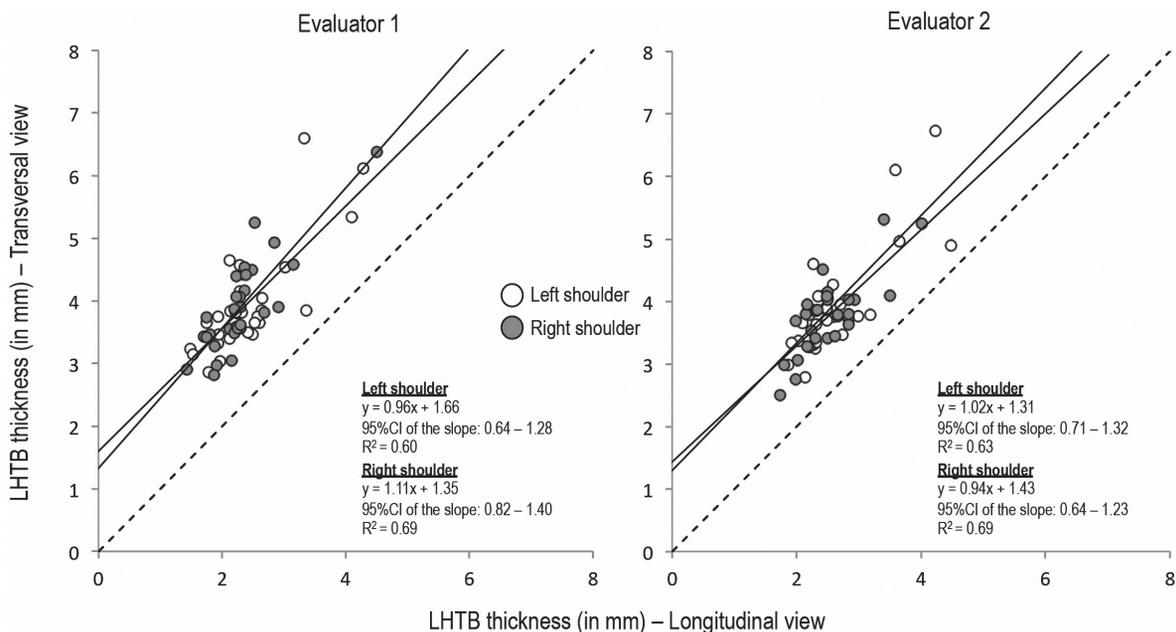

Fig. 2. Linear regression between longitudinal and transverse views of the long head of the biceps tendon thickness. *Dotted line* represents an absolute association between both measures. LHBT: long head of the biceps tendon.

et al. (14), the lack of experience of sonographers does not seem to negatively affect reliability.

The lower reliability for the width might be explained by the blurry aspect of the tendon boundaries in the transverse view, leading to a decrease in measurement precision (9). Difficulties faced when trying to identify the borders of the tendon accurately could be caused by the orientation of the probe (the angle at which the probe is held relative to the skin), since anisotropy can affect tendon appearance. The tendon imaging may also have been influenced by non-optimal US settings for this view (settings were optimized for thickness), which again will affect tendon appearance and decrease precision. Clinicians must be aware that measurement of LHBT width is associated with a more important measurement error than thickness. Therefore, based on reliability and MDC, thickness should be preferred over width when using just one measurement in clinics.

It is important to consider the MDC when evaluating change in a patient's status, since it can be used to determine whether the change is statistically and clinically meaningful. For example, the LHBT has been shown to be thickened in patients with supraspinatus full-thickness tear (15). The mean difference (0.61 mm) for LHBT thickness between patients with or without supraspinatus full-thickness tear is larger than the MDC for intra-rater measurement, and is thus outside the measurement error. Therefore, this measure could be used to evaluate change over time. However, the mean difference for LHBT thickness is smaller than the MDC for inter-rater measurement, which limits the use of this measure when performed by 2 different evaluators.

A systematic bias between measures of thickness using longitudinal and transverse views was observed, since the 2 measures were not centred on the dotted line, and the 95% CI for all 4 slopes included the expected systematic bias value of 1 (Fig. 2). The evaluation of thickness using transverse and longitudinal views, therefore, should be considered independently. While there was a statistical difference between these 2 thickness measures, the difference was constant (between 1.3 and 1.7 mm); showing that the difference is not due to a measurement error, but to a different placement of the probe on the LHBT leading to a different image of the same tendon.

The LHBT is a relatively easy tendon to image by ultrasound given its location on the humerus (superficial and anterior). This could contribute to the excellent reliability results obtained. The excellent reliability of most measures could also be attributed to the use of a standardized protocol for imaging, since we used pre-specified ultrasound parameters for all measurements. The protocol proposed in this study is simple and can be completed in less than 5 min; thus it can be included in the clinical evaluation of the shoulder. The main limitation of this study is that included participants were all asymptomatic. As reported by Collinger et al. (9), imaging healthy tendons can inflate reliability estimates that may not apply to degenerated tendons, which can be harder to image. Therefore, now that the reliability has been confirmed in healthy individuals, the next

step would be to examine the reliability and validity (sensitivity, specificity) of these measures in symptomatic shoulders since they will be of use in symptomatic populations in clinics.

In conclusion, quantitative measures of the LHBT using an ultrasound-imaging system are reliable in healthy shoulders. Based on the reliability and measurement errors, a mean of 2 measurements is recommended and thickness should be preferred over width, as the measurement of width is associated with lower reliability. Now that reliability has been shown in healthy shoulders, the next step will be to examine the validity and reliability of these quantitative measures in symptomatic shoulders.

REFERENCES

1. Ahrens PM, Boileau P. The long head of biceps and associated tendinopathy. J Bone Joint Surg Br 2007; 89: 1001–1009.
2. Walch G, Nove-Josserand L, Boileau P, Levigne C. Subluxations and dislocations of the tendon of the long head of the biceps. J Shoulder Elbow Surg 1998; 7: 100–108.
3. Chen CH, Hsu KY, Chen WJ, Shih CH. Incidence and severity of biceps long head tendon lesion in patients with complete rotator cuff tears. J Trauma 2005; 58: 1189–1193.
4. Murthi AM, Vosburgh CL, Neviaser TJ. The incidence of pathologic changes of the long head of the biceps tendon. J Shoulder Elbow Surg 2000; 9: 382–385.
5. Teefey SA, Hasan SA, Middleton WD, Patel M, Wright RW, Yamaguchi K. Ultrasonography of the rotator cuff. A comparison of ultrasonographic and arthroscopic findings in one hundred consecutive cases. J Bone Joint Surg Am 2000; 82: 498–504.
6. Armstrong A, Teefey SA, Wu T, Clark AM, Middleton WD, Yamaguchi K, et al. The efficacy of ultrasound in the diagnosis of long head of the biceps tendon pathology. J Shoulder Elbow Surg 2006; 15: 7–11.
7. Ahovuo J, Paavolainen P, Slatis P. Diagnostic value of sonography in lesions of the biceps tendon. Clin Orthop Relat Res 1986; 202: 184–188.
8. Skendzel JG, Jacobson JA, Carpenter JE, Miller BS. Long head of biceps brachii tendon evaluation: accuracy of preoperative ultrasound. Am J Roentgenol 2011; 197: 942–948.
9. Collinger JL, Gagnon D, Jacobson J, Impink BG, Boninger ML. Reliability of quantitative ultrasound measures of the biceps and supraspinatus tendons. Acad Radiol 2009; 16: 1424–1432.
10. Skou ST, Aalkjaer JM. Ultrasonographic measurement of patellar tendon thickness–a study of intra- and interobserver reliability. Clin Imaging 2013; 37: 934–937.
11. Portney LG, Watkins MP. Foundations of Clinical Research: Applications to 510 Practice. 3rd ed. Upper Saddle River, NJ: Pearson/Prentice Hall; 2008.
12. Fleiss JL. Reliability of measurement. The design and analysis of clinical experiments. Wiley Online Library. 2011 [accessed 2016 Jan 6]. Available from: http://dx.doi.org/10.1002/9781118032923.ch1.
13. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986; 1: 307–310.
14. Wallwork TL, Hides JA, Stanton WR. Intrarater and interrater reliability of assessment of lumbar multifidus muscle thickness using rehabilitative ultrasound imaging. J Orthop Sports Phys Ther 2007; 37: 608–612.
15. Chang KV, Chen WS, Wang TG, Hung CY, Chien KL. Quantitative ultrasound facilitates the exploration of morphological association of the long head biceps tendon with supraspinatus tendon full thickness tear. PLoS One 2014; 9: e113803.