



FACULDADE DE CIÊNCIAS DA NUTRIÇÃO E ALIMENTAÇÃO
UNIVERSIDADE DO PORTO

**Intra- and Inter-rater Reliability of Biceps Thickness Assessment by
Ultrasound in Elderly**

**Confiabilidade Intra- e Inter-individual na Avaliação da Espessura do Bíceps
por Ecografia em Idosos**

Ana Rita Carvalho Azevedo

Supervisor: João Pedro Pinho, MSc, RD

**Co-supervisors: Harriët Jager-Wittenaar, PhD, RD, and Maria Antónia
Vigário, MSc, RD**

Investigation Work

Bachelor Degree in Food Science

Faculty of Nutritional and Food Science of University of Porto

Porto, 2017

ABSTRACT AND KEY-WORDS

Introduction: This study aimed to assess intra- and inter-rater reliability of biceps thickness assessment by ultrasound (US) in elderly.

Methods: Thirty elderly (80% women), with a mean age of 82 years (standard deviation [SD] 6.3 years), and mean BMI of 26.7 kg/m² (SD 5.4 kg/m²), living in a nursing home or an adult day center, were included. To assess intra- and inter-rater procedural reliability, ultrasound measurements were performed by two raters considered “beginners” (R1, R2) on the biceps of the right arm. R1 repeated ultrasound measurements. To assess intra- and inter-rater measurement reliability, the first ultrasound image was analysed by three raters: R1, R2 and R3 who was considered experienced. All ultrasound measurements were performed with a BodyMetrix™ BX2000 device. Reliability was analysed by intraclass correlation coefficient (ICC), and was considered moderate to good if ICC=0.50-0.75 and good to excellent if ICC=0.75-1.00. Statistical significance was set at $p < 0.05$.

Results: ICC was 0.630 for intra-rater procedural reliability. ICC for inter-rater procedural reliability of image1 (R1) vs. image 2 (R2) was 0.622, and 0.534 for image 2 (R2) vs. image 3 (R1). ICCs for intra-rater measurement reliability of R1 and R2 were 0.865 and 0.766, respectively. For inter-rater measurement reliability, ICC of R1 vs. R2, R2 vs. R3, and R1 vs. R3 were 0.865, 0.800 and 0.815, respectively. Results were statistically significant ($p \leq 0.001$).

Conclusion: Procedural reliability of biceps thickness assessment in elderly is moderate to good, whereas measurement reliability is good to excellent. Increasing the level of experience may further improve procedural reliability.

Key-words Elderly, ultrasound measurement, biceps, reliability, muscle thickness.

RESUMO E PALAVRAS-CHAVE

Introdução: Este estudo pretende avaliar a confiabilidade intra- e inter-observador na avaliação da espessura do bíceps em idosos, por ecografia.

Métodos: Trinta idosos residentes em lares ou frequentadores de centro de dia (80% mulheres), com idade média de 82 anos (desvio-padrão [DP] de 6,3 anos), foram incluídos. A avaliação da confiabilidade do procedimento intra- e inter-observador, foram realizadas por dois observadores iniciantes (R1, R2) no bíceps do braço direito. O R1 repetiu as medições ecográficas. A Avaliação da confiabilidade da medição intra- e inter-observador, da primeira imagem ecográfica foi analisada por 3 observadores: R1, R2 e R3 (experiente). As medições foram realizadas com o aparelho BodyMetrix™ BX2000. A confiabilidade foi analisada pelo coeficiente de correlação intraclassa (ICC), (0,50-0,75 moderada a boa; 0,75-1,00 boa a excelente). Significado estatístico foi verificado quando $p < 0,05$.

Resultados: ICC= 0,630 para a confiabilidade do procedimento intra-observador. O ICC para a confiabilidade do procedimento inter-observador entre imagem 1 (R1) vs. imagem 2 (R2) foi de 0,622, e de 0,534 entre imagem 2 (R2) vs. imagem 3 (R1). Os ICCs para a confiabilidade da medição intra-observador no R1 e R2 foram 0,865 e 0,766, respetivamente. Para a confiabilidade da medição inter-observador, os ICCs entre R1 vs. R2, R2 vs. R3, e R1 vs. R3 foram 0,865, 0,800 e 0,815, respetivamente com $p \leq 0,001$.

Conclusão: A confiabilidade do procedimento na avaliação da espessura muscular em idosos é moderada a boa e a confiabilidade da medição é boa a excelente. O incremento de experiência pode melhorar a confiabilidade do procedimento.

Palavras-Chave: Idosos, medição ecográfica, bíceps, confiabilidade, espessura muscular.

CONTENT

ABSTRACT AND KEY-WORDS.....	i
RESUMO E PALAVRAS-CHAVE.....	iii
INTRODUCTION	1
AIM OF THE STUDY	3
METHODS AND PARTICIPANTS.....	3
RESULTS	8
DISCUSSION.....	9
CONCLUSION	13
ACKNOWLEDGMENTS.....	17
REFERENCES	19
APPENDICES.....	23

INTRODUCTION

Body composition assessment is an integral part of nutritional assessment, for example to diagnose malnutrition, and/or sarcopenia. Decreased fat-free mass (FFM) is an important criteria for malnutrition⁽¹⁾. According to the European Working Group on Sarcopenia in Older People (EWGSOP), sarcopenia is characterized by loss of muscle mass and muscle strength or performance^(2, 3). Besides identifying patients with low muscularity⁽⁴⁾, body composition assessment also allows to evaluate and monitor effects of interventions to improve nutritional status, like changes in dietary intake, exercise, and/or pharmacologic treatment.

There are various methods for body composition assessment, such as dual-energy X-ray absorptiometry (DEXA), computed tomography (CT), magnetic resonance imaging (MRI), bioelectrical impedance (BIA) and ultrasound (US)^(5, 6).

US is a valid⁽⁷⁾, bedside, simple, quick, portable, real-time method without ionizing radiation imaging, which allows not only to quantify muscle thickness but also to qualify the muscle (for example, assessment of fat infiltration or connective tissues). In contrast to DEXA and CT, US allows doing repeated measurements without any consequences to one's health. US is a method less expensive than CT and MRI and does not require high skills and experience as for CT and MRI^(1, 5, 8-10). Besides that, US is able to assess small muscles, for example biceps. Biceps muscle assessment would have clinical value, as nutrition deprivation is known to lead to a muscle loss mostly in upper body muscles, for example muscles in the arms⁽¹¹⁾. All these advantages make US a promising method in clinical and residential settings.

There are two types of ultrasound devices: Amplitude-Mode US (A-Mode US) and Brightness-Mode US (B-Mode US). Examples of A-mode US images

(BodyMetrix™ BX2000) and B-mode US images are available in Appendix 1.

One limitation associated to US (and not solved yet) is the absence of consensus on the best protocol that should be followed in ultrasound measurements⁽¹²⁾. Therefore, the exact measuring points of a specific muscle and the level of compression to apply are also not established⁽¹⁾. This causes variability in performance of US measurements, which results in poor reliability that includes both intra- and inter-rater reliability of US measurements. Another challenge is the difficulty to analyse US images, when determining muscle thickness. Adipose tissue and muscle tissue have similar acoustics impedances and, consequently, the boundary between these two tissues could be wrongly delineated. Therefore, it can decrease inter-rater reliability. Another factor that may influence inter-rater reliability is the experience of each rater in US^(1, 13).

Previous studies in which intra- and inter-reliability of muscle thickness was assessed US, were mostly performed in Intensive Care Unit patients^(4, 7, 14, 15). Only few studies were performed in older adults^(16, 17). Quadriceps was the most assessed muscle^(4, 14, 18, 19) and almost all studies used a B-mode US^(4, 14-17, 20-22). The majority of them included two raters^(14-18, 21, 23), and, in general, intra-reliability was higher than inter-reliability, despite of both being excellent^(4, 18, 22, 23). It is important to take into account that the excellent reliability was found in studies that assessed large muscles^(4, 18), that are easier to assess than small muscles, and in studies that had more than one rater with experience in ultrasound^(4, 18, 22), which could positively influence the results. Besides that, the different cut-offs used in studies could result in differences in evaluation of reliability. Only one study distinguished the reliability related to the way that procedure was followed, i.e. procedural reliability, and the reliability associated to the analysis of ultrasound images, i.e. measurement reliability⁽¹⁶⁾.

Therefore, in this study both “procedural reliability” and “measurement reliability” were addressed, as well as both intra- and inter-rater reliability.

AIM OF THE STUDY

This study aims to assess intra- and inter-rater procedural and measurement reliability of biceps muscle thickness assessment by US in elderly.

Hypotheses

- Intra-rater procedural reliability in biceps muscle thickness assessment by US in elderly is moderate;
- Inter-rater procedural reliability in biceps muscle thickness assessment by US in elderly is moderate;
- Intra- and inter-rater measurement reliability of muscle thickness by US in elderly is higher than its intra- and inter-rater procedural reliability.

METHODS AND PARTICIPANTS

Forty-one persons living in a nursing home (Vale S. Cosme or São Pedro de Bairro) or adult day center (Vale S.Cosme) were recruited for this observational cross-sectional study (7 men and 34 women). The following inclusion criteria were used: male or female; living in a nursing home or in an Adult Day Center since at least 1 month; age older than 65 years. The following exclusion criteria were used: being bedridden; taking muscle relaxants; having any scar on the arm that was measured due to any surgery or burn; having any neuromuscular disease; having unclear ultrasound images after imaging acquisition. Before participating, all participants were verbally informed about experimental procedures and they signed an informed consent, which also included the explanation about experimental procedures.

Sex and age were noted, however anonymity of participants was maintained.

Anthropometric measurements (weight and mid upper arm circumference [MUAC]) and US measurements (upper arm anterior-biceps) were performed in each participant. Due to lack of a stadiometer, the height of participants was retrieved from a clinical record. Body Mass Index (BMI) was calculated and was categorized according to the cut-offs defined by World Health Organization⁽²⁴⁾.

In this study three raters participated: two (rater 1 and rater 2) had “beginners’ level” of experience, and rater 3 was experienced in performing and analysing ultrasound images. Rater 1 performed anthropometric measurements, US measurements and US imaging analysis, rater 2 performed US measurement and US imaging analysis, and rater 3 performed imaging analysis.

Anthropometric measurements acquisition

Rater 1 performed all anthropometric measurements. To obtain anthropometric measurements of each participant, the right side of the body was used and anthropometric landmarks were previously marked, according the techniques of International Standards of Anthropometric Assessment⁽²⁵⁾. During anthropometric measurements, subjects were standing and relaxed. For this procedure, a weighing scale (SECA, model 761), large sliding caliper, segmometer and anthropometric tape were used. The anatomical landmarks which were used to obtain arm circumference were acromiale point, radiale point and mid-acromiale-radiale point. To ensure that anthropometric measurements and ultrasound measurements were performed at the same location, these sites were previously marked by a surgery pen.

Biceps measurements

In imaging acquisition, cross-sectional images of the biceps were obtained. In

imaging analysis, the largest thicknesses of the biceps (in millimeters) was assessed.

Imaging Acquisition

Images were acquired in the following order: rater 1 ("Proc1"), rater 2 ("Proc2") and once again rater 1 ("Proc3"). Proc1 and Proc3 were performed with a time interval of 5 minutes. Rater 1 and 2 measured all participants and were blinded to the other rater's readings. The device BodyMetrix™ BX2000 (IntelaMetrix), a 2.5 MHz A-mode ultrasound, coupled with Body View Professional Software version 1.0.6.7425 (IntelaMetrix) was used for imaging acquisition and imaging analysis.

During the imaging acquisition, subjects were standing and relaxed, and the right arm in supine position was used, in fixed position. Conducting gel was placed in the ultrasound transducer head and in the measurement site of the participant, to reduce friction and allow the transducer to easily move on the skin. Then, scanning started from 1 cm above the mid-acromiale-radiale anterior point until distal insertion of biceps, during 5 seconds. The protocol developed for the imaging acquisition can be found in Appendix 2.

Imaging Analysis

Imaging analysis was performed by rater 1, rater 2 and rater 3. Rater 1 analysed images generated from "Proc1" and "Proc3". After 7 days⁽²⁶⁾, rater 1 analysed again images generated from "Proc1". The period of 7 days was considered enough to not remember the first series of analysis. Therefore, images analysed by rater 1 were labelled as "Proc1_R1_M1", "Proc3_R1" and "Proc1_R1_M2". Rater 2 analysed images from "Proc2" acquired by himself, images from "Proc1" performed by rater 1, and 7 days after, rater 2 re-analysed "Proc1". Images analysed by rater 2 were labelled as "Proc2_R2", "Proc1_R2_M1" and "Proc1_R2_M2". Rater 3 only once analysed "Proc1" acquired by rater 1. Images analysed by rater 3 were labelled as "Proc1_R3".

Muscle thickness (in millimeter; mm) was determined from the interface between adipose tissue and muscle tissue (fascia) to the interface between muscle tissue and bone tissue. Each rater measured the largest biceps thickness in each US image three times, and then calculated the mean of these three biceps thicknesses.

Statistical Analysis

A calculation was performed to assess in which sample size the statistical tests would have sufficient “power”, i.e. if the study findings are valid to detect a difference⁽²⁷⁾. For this calculation we used a sample size calculator developed by the Clinical & Translational Science Institute⁽²⁸⁾, based on a correlation coefficient of $r=0.5$ ⁽²⁹⁾, type I error $\alpha=0.05$ and a type II error $\beta=0.20$ ⁽²⁷⁾. The calculation resulted in a needed number of 29 participants. Data were entered in a database [Excel version 2010]. Before the start of the analysis, data were checked on errors of transcription. Normality of data was assessed by Shapiro-Wilk test⁽³⁰⁾. Descriptive statistics were used, in which categorical variables are presented as numbers and percentages, continuous variables are presented as mean and standard deviation (SD). For statistical analysis, eight pairs of variables were defined, as shown in table 1.

To assess the level of agreement between the different imaging analyses, Pearson Correlation Coefficient (r) was used. Correlation was considered very poor if $r=0$ to 0.25 , poor if $r=0.25$ to 0.5 , moderate if $r=0.5$ to 0.75 , strong if $r=0.75$ to 0.9 , and very strong if $r=0.9$ to 1 ⁽³¹⁾. Reliability between different imaging analyses was assessed by Intraclass Correlation Coefficient (ICC). For intra-rater reliability, the one-way random model (absolute agreement) was used. For inter-rater reliability, the two-way mixed model (absolute agreement) was used. Reliability was considered non-existent if $ICC=0.00$ to 0.25 , fair if $ICC=0.25$ to 0.50 , moderate to good if $ICC=0.50$ to 0.75 and

good to excellent if ICC=0.75 to 1.00⁽³²⁾. Bland-Altman plots were constructed to visualize differences between imaging analyses, and any systematic bias or outliers.

Statistical significance was set at $p < 0.05$. All data were analysed through by SPSS version 24.0 (SPSS Inc., an IBM Company, USA) and Bland-Altman plots were constructed in statistical analysis software XLSTAT version 2017.1.

Table 1. Pairs defined for statistical analysis of intra- and interrater reliability

Pairs				
MEASUREMENT RELIABILITY	Inter- Rater	Pair 1	Proc1_R1_M1	Proc1_R2_M1
		Pair 2	Proc1_R1_M1	Proc1_R3
		Pair 3	Proc1_R2_M1	Proc1_R3
	Intra- Rater	Pair 4	Proc1_R1_M1	Proc1_R1_M2
		Pair 5	Proc1_R2_M1	Proc1_R2_M2
PROCEDURAL RELIABILITY	Inter- Rater	Pair 6	Proc1_R1_M1	Proc2_R2
		Pair 7	Proc3_R1	Proc2_R2
	Intra- Rater	Pair 8	Proc1_R1_M1	Proc3_R1

Proc1: Procedure 1 acquired by rater 1; Proc2: Procedure 2 acquired by rater 2; Proc3: Procedure 3 acquired by rater 1; R1: Analysis done by rater 1; R2: Analysis done by rater 2; R3: Analysis done by rater 3; M1: Moment 1 of analysis; M2: Moment 2 of analysis. Proc1_R1_M1: Procedure 1 analysed by rater 1 in moment 1; Proc1_R2_M1: Procedure 1 analysed by rater 2 in moment 1; Proc1_R3: Procedure 1 analysed by rater 3; Proc1_R1_M2: Procedure 1 analysed by rater 1 in moment 2; Proc1_R2_M2: Procedure 1 analysed by rater 2 in moment 2; Proc2_R2: Procedure 2 analysed by rater 2; Proc3_R1: Procedure 3 analysed by rater 1.

Ethics

The study was performed according to the Declaration of Helsinki - ethical principles for medical research involving human subjects⁽³³⁾, and the study was approved by the Ethics Committee of the University of Porto.

RESULTS

Characterization of the study sample

From the initial 41 participants, 11 participants were excluded because in 10 participants ultrasound images were not clear and 1 had a neuromuscular disease (amyotrophic lateral sclerosis). Therefore, the final sample size was 30 older adults, with a mean age of 82 years (SD 6.3 years), of which 80% were women (n=24). Mean BMI was 26.7 kg/m² (SD 5.4 kg/m²). Finally, the mean upper arm circumference of all 30 older adults was 28.0 cm with a SD of 4.1 cm. Characteristics of the study sample can be found in Appendix 3. Mean BMI and MUAC of excluded people did not differ from the people included in the analysis.

Reliability

In Appendix 4, the mean of the three largest biceps thicknesses of each US image, by each rater, is presented. All of them have normal distribution ($p>0.05$) and data were complete.

Measurement Reliability

For the assessment of inter-rater measurement reliability, pair 1, pair 2 and pair 3 were used. To assess intra-rater measurement reliability, pairs 4 and 5 were used. Results are presented in table 2. In Bland-Altman plots, in pairs 1 and 4 it was found a negative correlation between the mean of the largest biceps thicknesses compared in each pair and the difference between these two largest biceps thicknesses. In the other Bland-Altman plots of the others pairs, a positive correlation was verified between these two variables. These plots can be found in Appendix 5.

Procedural Reliability

For the assessment of inter-rater procedural reliability, pairs 6 and 7 were used

and to assess intra-rater procedural reliability pair 8 was used. Results are also presented in table 2. In all Bland-Altman plots it was verified a positive correlation between the mean of the largest biceps thicknesses compared in each pair and the difference between these two largest biceps thicknesses. These plots can also be found in Appendix 5.

DISCUSSION

The results of this study indicate that assessment of biceps muscle in older adults can be performed reliably. We found a moderate to good intra- and inter-rater procedural reliability and a good to excellent intra- and inter-rater measurement reliability. In addition, measurement reliability of raters considered beginners was not different from measurement reliability of the rater considered experienced in imaging analysis.

As hypothesized, both intra- and inter-rater procedural reliability in muscle thickness assessment of elderly were moderate to good. However, intra- and inter-rater measurement reliability of muscle thickness by US were higher than intra- and inter-rater procedural reliability of muscle thickness by US. This finding was also in line with our hypothesis, since repeating the full procedure for performing an ultrasound measurement requires more handling than other steps in the imaging analysis and, consequently, it could have a higher error associated to the rater.

This is the first study that assessed reliability of biceps muscle thickness assessment using BodyMetrix™ BX2000 and therefore it is not possible to compare this study with other studies. Thus far, BodyMetrix™ BX2000 was mostly used to quantify FM through prediction equations predefined in software⁽³⁴⁻³⁷⁾. In some studies this device was also used to estimate whole body FFM from multiple point measurements^(35, 36).

Table 2. Results of measurement and procedural reliability.

		r		ICC		Bland-Altman Plot			
		p	r value	p	ICC value	Bias	Upper LOA	Lower LOA	R ²
Pair 1	Proc1_R1_M1	<0.001	0.884	<0.001	0.865	1.044	5.335	-3.246	0.0016
	Proc1_R2_M1								
Pair 2	Proc1_R1_M1	<0.001	0.820	<0.001	0.815	0.571	5.712	-4.570	0.0193
	Proc1_R3								
Pair 3	Proc1_R2_M1	<0.001	0.804	<0.001	0.800	-0.473	4.962	-5.909	0.0271
	Proc1_R3								
Pair 4	Proc1_R1_M1	<0.001	0.888	<0.001	0.865	1.071	5.351	-3.209	0.0067
	Proc1_R1_M2								
Pair 5	Proc1_R2_M1	<0.001	0.843	<0.001	0.766	-1.535	3.324	-6.393	0.1619
	Proc1_R2_M2								
Pair 6	Proc1_R1_M1	<0.001	0.638	<0.001	0.622	-0.648	6.310	-7.605	0.0770
	Proc2_R2								
Pair 7	Proc3_R1	0.003	0.529	0.001	0.534	-0.391	6.657	-7.439	0.0032
	Proc2_R2								
Pair 8	Proc1_R1_M1	<0.001	0.633	<0.001	0.630	-0.257	6.825	-7.338	0.0480
	Proc3_R1								

r: Pearson Correlation Coefficient; ICC: Intraclass Correlation Coefficient; LOA: Limits of agreement; Proc1_R1_M1: Procedure 1 analysed by rater 1 in moment 1; Proc1_R2_M1: Procedure 1 analysed by rater 2 in moment 1; Proc1_R3: Procedure 1 analysed by rater 3; Proc1_R1_M2: Procedure 1 analysed by rater 1 in moment 2; Proc1_R2_M2: Procedure 1 analysed by rater 2 in moment 2; Proc2_R2: Procedure 2 analysed by rater 2; Proc3_R1: Procedure 3 analysed by rater 1; R²: regression line.

Only few studies reported on inter-rater measurement reliability of ultrasound.

In one study, which assessed thigh muscles of healthy subjects in bed-rest position by two raters with some training in US, the results were slightly better (ICC=0.963-0.991)⁽²¹⁾ than the ones obtained by our two raters considered “beginners” (ICC=0.865). One possible reason for our lower reliability found may be existence of differences in methodology used in each study: while in that study the participants were healthy adults that were in bed-rest position during imaging acquisition, in our study the participants were older adults that were standing during imaging acquisition. Bed-rest position requires less mobilization to perform the measurements than standing position. Besides that, older adults have more difficulties in maintaining requested position than younger adults. Moreover, with ageing, muscle quality gets worse, which may increase fat infiltration in the muscle⁽¹⁶⁾, which may hinder image quality and measurement reliability.

In another study, which evaluated rectus femoris muscle of patients with critical illness, the reliability found between US users at beginners level were similar to ours (ICC=0.851-0.968)⁽⁷⁾. With regard to measurement reliability between “beginner” rater and “experienced” rater, one study that assessed lumbar multifidus muscle (a deep muscle located along the back of the spine) in community-dwelling older adults in supine position by two raters (one considered beginner and the other considered experienced in US) found a good to excellent reliability (ICC=0.91-0.93)⁽¹⁶⁾, also slightly higher than ours (ICC=0.815 and ICC=0.800). One possible reason for our results being lower may be the difference between participants’ position in these studies: in our study participants were standing and not in supine position, being more difficult to maintain the requested position, which can negatively affect the results.

No studies on intra-rater measurement reliability, in any kind of muscle

assessment by ultrasound, were found, which hinders comparison of our results with others. With regard to procedural reliability of ultrasound, it is important to emphasize that, in our study, imaging acquisition was only performed by raters considered “beginners”, which could have negatively affected reliability. Considering inter- and intra-rater procedural reliability, three studies were found, which included raters considered “not experienced”, with better results than ours (Inter-Rater: ICC=0.71⁽²³⁾, ICC=0.995⁽¹⁵⁾ and ICC=0.70-0.99⁽¹⁷⁾; Intra-Rater: ICC=0.97⁽²³⁾, ICC=0.949⁽¹⁵⁾ and ICC=0.69-0.99⁽¹⁷⁾). Differences in methodology and sample may have caused this discrepancy between the results of other studies and ours. One study was performed in a young sample (mean age 18.8 years [SD 2.5 years]), and younger muscles may be easier to assess⁽²³⁾. Another study performed US measurements in muscles bigger than biceps muscles⁽¹⁷⁾, which can also originate better results than ours.

In two Bland-Altman plots, i.e. in pair 1 (inter-rater measurement reliability) and in pair 4 (intra-rater measurement reliability), a negative correlation between the average and the differences were found, which indicates that the higher the average, the smaller the difference. This might suggest that reliability might be better in thicker muscles.

This research can draw multiple implications for practice. The fact that both measurement and procedural reliability were assessed allows to understand that US can be performed with a moderate procedural reliability and an excellent measurement reliability by raters considered not experienced in US, for example by nutritionists. Since US is a portable, quick, safe⁽⁵⁾, and not expensive method, it has additional value for nutritional assessment, to diagnose, evaluate and monitor malnutrition and or sarcopenia. However, the introductory course “*Body Composition Assessment: An Introduction to Computed Tomography and Ultrasound*”, with 1.5 hours of practising

with BodyMetrix™ BX2000, attended by the raters with “beginners level” prior to this study, was not sufficient to achieve an excellent reliability. More training could be beneficial to improve procedural reliability. In addition, a protocol was created for this study, with the aim to standardize the measurements as much as possible. This protocol can be further used by other researchers in the future.

This study also has some limitations. The use of different gels during imaging acquisition, due to lack of resources, could be considered a limitation. Some of the gels were not compatible with the device used and, consequently, it caused differences in friction between the device and the skin, and differences in the compression used by raters during imaging acquisition. In addition, the study design was limited by the fact that rater 3, who was experienced in imaging analysis, did not perform imaging acquisition. Therefore, we could not compare imaging acquisitions performed between raters with “beginners” level and the “experienced” rater, which hindered us to understand if different levels of experience in US affect procedural reliability. Finally, 11 of 41 initial participants could not be included in this study. In 10 of the excluded people, ultrasound images were not clear. In the literature, the existence of poor images is rarely reported, however it could be due to the difficulty to find a reason for that. However, it cannot be ruled out that the poor quality of the image might be related to the image procedure.

CONCLUSION

The results of this study show a moderate to good intra- and inter-rater procedural reliability and good to excellent intra- and inter-rater measurement reliability. This suggests that US can be a bedside tool for biceps muscle thickness assessment in elderly even with raters with low skills in US method. Procedural reliability was lower

than measurement reliability, probably because raters that performed imaging acquisition only had a small training.

ACKNOWLEDGMENTS

The end of one more journey has come, where many different people contributed in such a way that I eventually reached the top of this mountain. All I have to do now is to be thankful towards them all, with much joy.

To my supervisor João Pinho, MSc, RD, that helped me to achieve this goal with all his dynamism and readiness.

To my co-supervisors, Harriët Jager-Wittenaar, PhD, RD, and Maria Antónia Vigário, MSc, RD, for all the support borrowed during the accomplishment of my work.

To Luiz da Silva, PhD, for sending me device BodyMetrix before the beginning of my work, allowing the execution of this work without problems.

To Thiago Silva, MSc, MD, for all his suggestions in the beginning of the execution of my work, contributing to improve its quality.

To Paula Rocha, BSc, Joana Silveira, MSc, and Willemke Nijholt, MSc, for helping me in gathering and or analysing the data, contributing to a successful work.

To my parents and sister for the effort, patience and affection. For all the good they do for me since forever.

And last but not least, to my Luís, for being my safe haven during this journey and for always believing in me. Far but near.

REFERENCES

1. Earthman CP. Body Composition Tools for Assessment of Adult Malnutrition at the Bedside: A Tutorial on Research Considerations and Clinical Applications. *JPEN Journal of parenteral and enteral nutrition*. 2015; 39(7):787-822.
2. Cruz-Jentoft AJ, Baeyens JP, Bauer JM, Boirie Y, Cederholm T, Landi F, et al. Sarcopenia: European consensus on definition and diagnosis: Report of the European Working Group on Sarcopenia in Older People. *Age Ageing*. 2010; 39(4):412-23.
3. Woo J, Leung J, Morley JE. Defining sarcopenia in terms of incident adverse outcomes. *Journal of the American Medical Directors Association*. 2015; 16(3):247-52.
4. Paris MT, Mourtzakis M, Day A, Leung R, Watharkar S, Kozar R, et al. Validation of Bedside Ultrasound of Muscle Layer Thickness of the Quadriceps in the Critically Ill Patient (VALIDUM Study). *JPEN Journal of parenteral and enteral nutrition*. 2017; 41(2):171-80.
5. Prado CMM, Heymsfield SB. Lean Tissue Imaging: A New Era for Nutritional Assessment and Intervention. *JPEN Journal of parenteral and enteral nutrition*. 2014; 38(8):940-53.
6. Siervo M, Jebb SA. Body composition assessment: theory into practice: introduction of multicompartiment models. *IEEE engineering in medicine and biology magazine : the quarterly magazine of the Engineering in Medicine & Biology Society*. 2010; 29(1):48-59.
7. Sarwal A, Parry SM, Berry MJ, Hsu FC, Lewis MT, Justus NW, et al. Interobserver Reliability of Quantitative Muscle Sonographic Analysis in the Critically Ill Population. *Journal of ultrasound in medicine : official journal of the American Institute of Ultrasound in Medicine*. 2015; 34(7):1191-200.
8. Pillen S, van Alfen N. Skeletal muscle ultrasound. *Neurol Res*. 2011; 33(10):1016-24.
9. C Toomey KM, S Leahy and P Jakeman. Technical considerations for accurate measurement of subcutaneous adipose tissue thickness using B-mode ultrasound. *Ultrasound*. 2011; 19
10. W. Nijholt I-JA, G. Adajar et al. Ultrasound assessment of muscle thickness and muscle cross-sectional area: a reliability study. In: Peter Hogg RH, Carst Buissink, editor. *OPTIMAX 2016-Optimising image quality for medical imaging*. University of Salford, Salford, Greater Manchester, UK; 2017. p. 71-82.
11. Ottery FJ-W. PG-SGA/Pt-Global Platform. 2014. Disponível em: <http://pt-global.org/>.
12. Levi M, Teigen B, RD, LD, Adam J, Kuchnia, MS, RD, LD, CNSC1,, Marina Mourtzakis P, and Carrie P. Earthman, PhD, RD, LD1. The Use of Technology for Estimating Body Composition: Strengths and Weaknesses of Common Modalities in a Clinical Setting. *Nutrition in Clinical Practice*. 2016; 20
13. Gellhorn AC, Carlson MJ. Inter-rater, intra-rater, and inter-machine reliability of quantitative ultrasound measurements of the patellar tendon. *Ultrasound in medicine & biology*. 2013; 39(5):791-6.
14. Fizez T, Hendrickx A, Van Herpe T, Vlasselaers D, Desmet L, Van den Berghe G, et al. An Analysis of Reliability and Accuracy of Muscle Thickness

Ultrasonography in Critically Ill Children and Adults. *JPEN Journal of parenteral and enteral nutrition*. 2016; 40(7):944-9.

15. Hadda V, Kumar R, Dhungana A, Khan MA, Madan K, Khilnani GC. Inter- and intra-observer variability of ultrasonographic arm muscle thickness measurement by critical care physicians. *Journal of postgraduate medicine*. 2017

16. Sions JM, Velasco TO, Teyhen DS, Hicks GE. Ultrasound Imaging: Intra- and Inter-Examiner Reliability for Multifidus Muscle Thickness Assessment in Adults Aged 60–85 Years versus Younger Adults. *The Journal of orthopaedic and sports physical therapy*. 2014; 44(6):425-34.

17. Cho KH, Lee HJ, Lee WH. Reliability of rehabilitative ultrasound imaging for the medial gastrocnemius muscle in poststroke patients. *Clinical physiology and functional imaging*. 2014; 34(1):26-31.

18. Tillquist M, Kutsogiannis DJ, Wischmeyer PE, Kummerlen C, Leung R, Stollery D, et al. Bedside Ultrasound Is a Practical and Reliable Measurement Tool for Assessing Quadriceps Muscle Layer Thickness. *JPEN Journal of parenteral and enteral nutrition*. 2014; 38(7):886-90.

19. Blazeovich AJ, Gill ND, Zhou S. Intra- and intermuscular variation in human quadriceps femoris architecture assessed in vivo. *Journal of anatomy*. 2006; 209(3):289-310.

20. Reeves ND, Maganaris CN, Narici MV. Ultrasonographic assessment of human skeletal muscle size. *European journal of applied physiology*. 2004; 91(1):116-8.

21. Scott JM, Martin DS, Ploutz-Snyder R, Caine T, Matz T, Arzeno NM, et al. Reliability and validity of panoramic ultrasound for muscle quantification. *Ultrasound in medicine & biology*. 2012; 38(9):1656-61.

22. Overas CK, Myhrvold BL, Rosok G, Magnesen E. Musculoskeletal diagnostic ultrasound imaging for thickness measurement of four principal muscles of the cervical spine -a reliability and agreement study. *Chiropractic & manual therapies*. 2017; 25:2.

23. Watson T, McPherson S, Fleeman S. Ultrasound measurement of transversus abdominis during loaded, functional tasks in asymptomatic individuals: rater reliability. *PM & R : the journal of injury, function, and rehabilitation*. 2011; 3(8):697-705.

24. Organization WH. Body Mass Index - BMI. 2017. Disponível em: <http://www.euro.who.int/en/health-topics/disease-prevention/nutrition/a-healthy-lifestyle/body-mass-index-bmi>.

25. Stewart A, Marfell-Jones M, International Society for Advancement of K. International standards for anthropometric assessment. Lower Hutt, New Zealand: International Society for the Advancement of Kinanthropometry; 2011.

26. Dunn LM, Dunn LM, Bulheller S, Häcker H. Peabody picture vocabulary test. American Guidance Service Circle Pines, MN; 1965.

27. Fitzner K, Heckinger E. Sample size calculation and power analysis: a quick review. *The Diabetes educator*. 2010; 36(5):701-7.

28. Kohn M JM, Senyak J. Sample Size Calculator. Clinical & Translational Science Institute. [atualizado em: 2017]. Disponível em: <http://www.sample-size.net/correlation-sample-size>.

29. Strasser EM, Draskovits T, Praschak M, Quittan M, Graf A. Association between ultrasound measurements of muscle thickness, pennation angle, echogenicity and skeletal muscle strength in the elderly. *Age (Dordrecht, Netherlands)*. 2013; 35(6):2377-88.

30. Ghasemi A, Zahediasl S. Normality Tests for Statistical Analysis: A Guide for Non-Statisticians. *International Journal of Endocrinology and Metabolism*. 2012; 10(2):486-89.
31. Mukaka MM. A guide to appropriate use of Correlation coefficient in medical research. *Malawi Medical Journal : The Journal of Medical Association of Malawi*. 2012; 24(3):69-71.
32. Portney LG, Watkins MP. *Foundations of clinical research: applications to practice*. Prentice Hall Upper Saddle River, NJ; 2000.
33. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *The Journal of the American College of Dentists*. 2014; 81(3):14-8.
34. Wagner DR, Cain DL, Clark NW. Validity and Reliability of A-Mode Ultrasound for Body Composition Assessment of NCAA Division I Athletes. *PLoS One*. 2016; 11(4):e0153146.
35. Smith-Ryan AE, Fultz SN, Melvin MN, Wingfield HL, Woessner MN. Reproducibility and validity of A-mode ultrasound for body composition measurement and classification in overweight and obese men and women. *PLoS One*. 2014; 9(3):e91750.
36. Schoenfeld BJ, Aragon AA, Moon J, Krieger JW, Tiryaki-Sonmez G. Comparison of amplitude-mode ultrasound versus air displacement plethysmography for assessing body composition changes following participation in a structured weight-loss programme in women. *Clinical physiology and functional imaging*. 2016
37. Johnson K, Miller B, McLain TA, Gibson AL, Otterstetter R. Bioelectrical impedance and ultrasound to assess body composition in college-aged adults. *Journal of Advanced Nutrition and Human Metabolism*. 2016; 2

APPENDICES

APPENDICES

Appendix 1 – A-Mode and B-mode ultrasound images.....	25
Appendix 2 – Protocol for Imaging Acquisition	27
Appendix 3 – Descriptive statistics of sociodemographic variables (n=30)	29
Appendix 4 – Biceps thickness* per measurement (mm).....	31
Appendix 5 – Bland-Altman Plots	33

Appendix 1 – A-Mode and B-mode ultrasound images

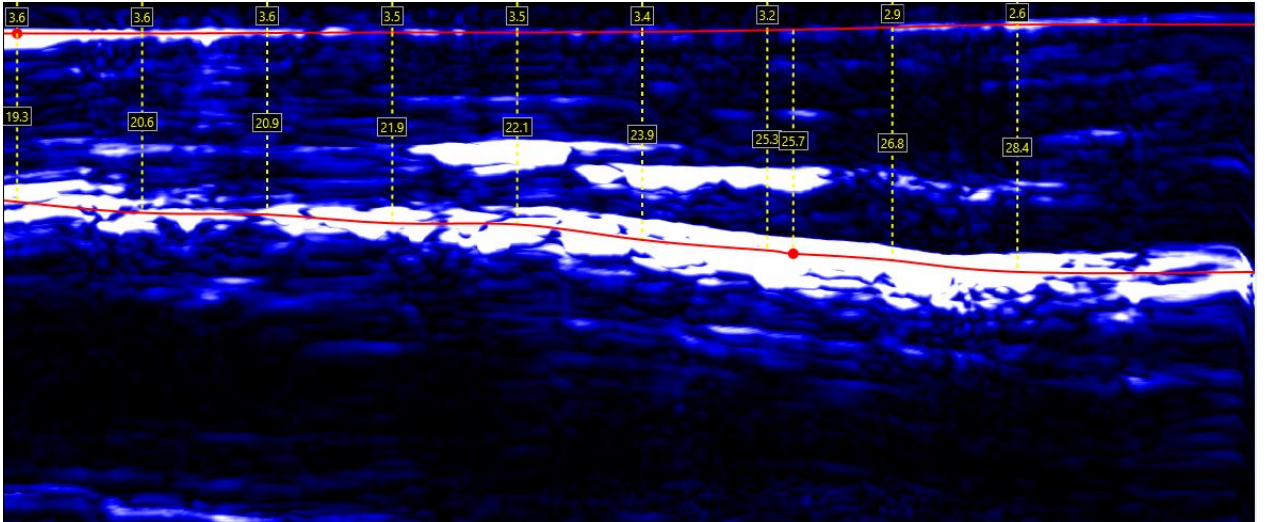


Figure 1. A-Mode ultrasound cross-sectional image of upper arm anterior- biceps.



Figure 2. B-Mode ultrasound cross-sectional image of rectus femoris.

Appendix 2 – Protocol for Imaging Acquisition

1. Put the participant standing and relaxed, advising him that imaging acquisition will be done in the right upper arm anterior zone of his body;
2. Ask to participant to put right arm relaxed and in supination;
3. Take the transducer of device BodyMetrix™ BX2000 (IntelaMetrix), 2,5MHz, A-mode ultrasound, and put conducting gel in the head of the transducer (to reduce friction and allow the transducer to easily move on the skin);



Figure 1. Placing conducting gel on the head of the transducer

4. Support the weight of the right arm of the participant with the hand that is not taking the transducer (with the aim to maintain the arm of the participant in a fixed position);
5. With the head of the transducer, spread gel in right upper arm anterior zone of participant;



Figure 2. Spreading gel in right upper arm anterior zone of participant

6. To obtain a cross-sectional image, orient the transducer transversally to the longitudinal axis of the arm, forming a 90° angle in relation to the skin surface;



Figure 3. Correct orientation of the transducer in relation to skin surface

7. Applying minimal pressure to the transducer, start scanning from 1 cm above the *mid-acromiale-radiale* point (previously marked) until distal insertion of biceps, during 5 seconds.

Appendix 3 – Descriptive statistics of sociodemographic variables (n=30)

Age (y)		Mean	82
		SD	6.3
Sex	Men	N	6
		Percentage (%)	20.0
	Women	N	24
		Percentage (%)	80.0
Weight (kg)		Mean	64.8
		SD	13.2
Height (m)		Mean	1.56
		SD	0.06
BMI (kg/m²)		Mean	26.7
		SD	5.4
BMI Categories (kg/m²)	Underweight (BMI <18.5 kg/m²)	N	1
		Percentage (%)	3.3
	Normal Weight (BMI 18.5-24.9 kg/m²)	N	12
		Percentage (%)	40.0
	Overweight (BMI 25.0-29.9 kg/m²)	N	9
		Percentage (%)	30.0
	Obesity (BMI ≥30.0 kg/m²)	N	8
		Percentage (%)	26.7
MUAC (cm)		Mean	28.0
		SD	4.1

SD: Standard Deviation; BMI: Body Mass Index; MUAC: Mid upper arm circumference.

Appendix 4 – Biceps thickness* per measurement (mm)

Personal ID	Proc1_R1_M1	Proc1_R2_M1	Proc1_R3	Proc1_R1_M2	Proc1_R2_M2	Proc2_R2	Proc3_R1
1935F30	19.6	20.5	18.5	18.2	19.8	23.3	19.3
1941F31	24.8	24.5	24.1	22.9	25.3	25.4	24.8
1938F33	23.4	24.9	24.3	20.4	24.5	27.0	26.9
1934M36	31.9	32.6	30.3	32.6	32.2	29.4	30.9
1933M48	22.7	17.7	20.6	18.7	26.5	24.4	23.1
1929F49	24.2	25.5	23.4	25.4	25.8	28.9	24.4
1923M50	25.6	27.6	25.3	28.1	27.9	29.7	28.8
1936M51	35.8	28.9	31.7	29.8	29.0	25.6	23.5
1929F54	25.5	24.3	23.7	25.7	24.3	25.2	20.3
1938F55	22.1	22.6	22.8	23.4	22.7	23.6	26.1
1933F56	24.5	23.6	23.9	25.8	25.1	30.5	26.7
1930F71	20.5	19.7	29.9	19.4	19.9	20.2	28.8
1940F73	23.5	21.0	21.4	22.2	23.1	24.8	23.3
1996F74	16.2	11.1	15.9	11.0	19.1	15.4	19.2
1932F75	18.2	18.2	17.6	18.8	18.4	20.4	20.1
1939F76	16.4	17.0	16.2	17.2	16.8	16.9	17.1
1996F77	18.3	19.1	22.1	18.1	24.6	24.4	20.6
1944F78	24.5	22.0	23.5	24.2	23.9	22.3	26.7
1938F79	19.8	20.2	22.0	21.7	21.5	25.4	21.5
1933M80	26.4	21.9	26.2	20.6	27.2	22.6	24.4
1928F81	24.2	21.3	20.8	21.6	23.4	21.7	25.4
1926M82	31.8	30.9	30.9	30.4	31.4	29.1	31.3
1923F83	21.8	22.0	22.6	22.3	22.6	23.8	21.4
1938F84	25.7	26.3	24.6	25.6	25.8	26.3	22.2
1945F86	15.8	13.0	16.2	13.5	18.3	19.6	17.0
1924F88	25.2	22.7	19.3	23.5	24.2	22.2	21.6
1929F90	21.5	21.4	20.7	18.9	22.2	23.6	21.0
1941F91	23.0	21.4	19.4	22.5	22.2	23.6	22.9
1932F93	24.2	24.0	22.9	22.9	24.1	20.2	28.7
1941F94	25.1	24.9	24.2	24.4	25.1	26.1	21.9

* Mean of three values

Proc1_R1_M1: Procedure 1 analysed by rater 1 in moment 1;

Proc1_R2_M1: Procedure 1 analysed by rater 2 in moment 1;

Proc1_R3: Procedure 1 analysed by rater 3;

Proc1_R1_M2: Procedure 1 analysed by rater 1 in moment 2;

Proc1_R2_M2: Procedure 1 analysed by rater 2 in moment 2;

Proc2_R2: Procedure 2 analysed by rater 2;

Proc3_R1: Procedure 3 analysed by rater 1.

Appendix 5 – Bland-Altman Plots

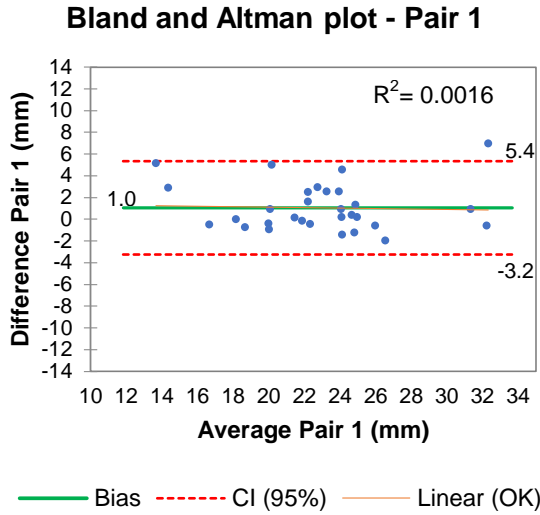


Figure 1. Bland-Altman analysis for biceps muscle thickness between Proc1_R1_M1 and Proc1_R2_M1 (inter-rater measurement reliability).

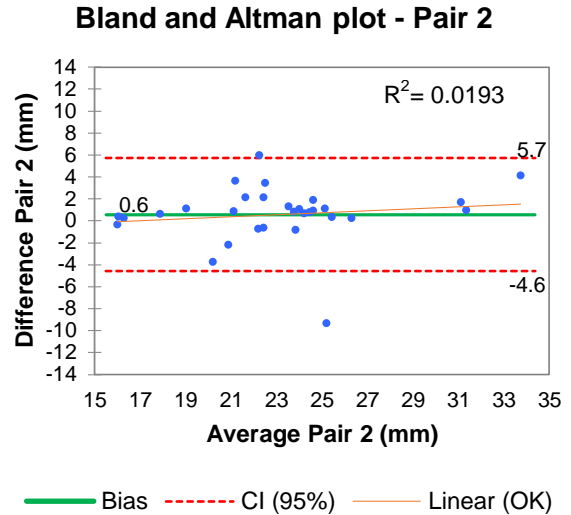


Figure 2. Bland-Altman analysis for biceps muscle thickness between Proc1_R1_M1 and Proc1_R3 (inter-rater measurement reliability).

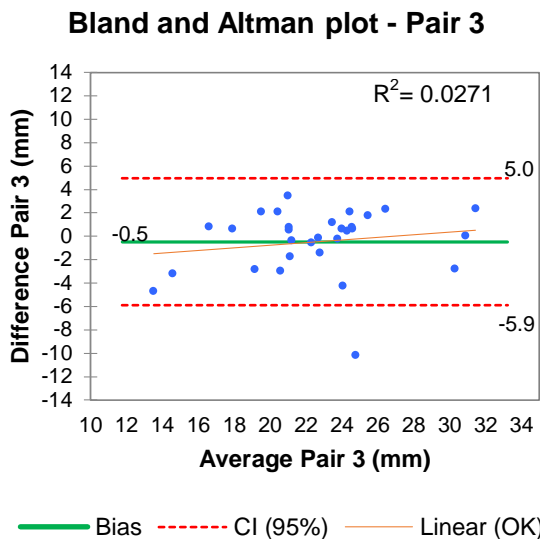


Figure 3. Bland-Altman analysis for biceps muscle thickness between Proc1_R2_M1 and Proc1_R3 (inter-rater measurement reliability).

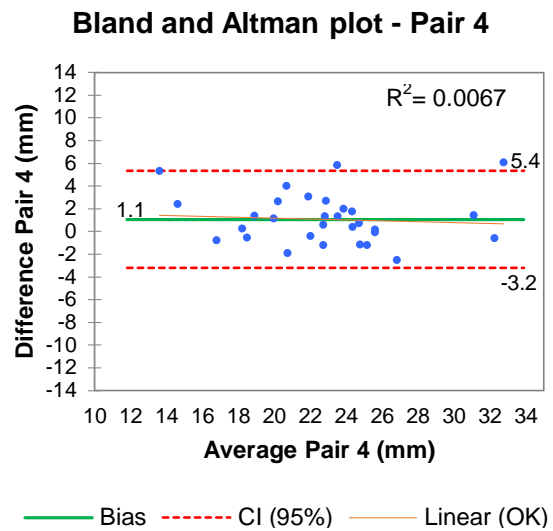


Figure 4. Bland-Altman analysis for biceps muscle thickness between Proc1_R1_M1 and Proc1_R1_M2 (intra-rater measurement reliability).

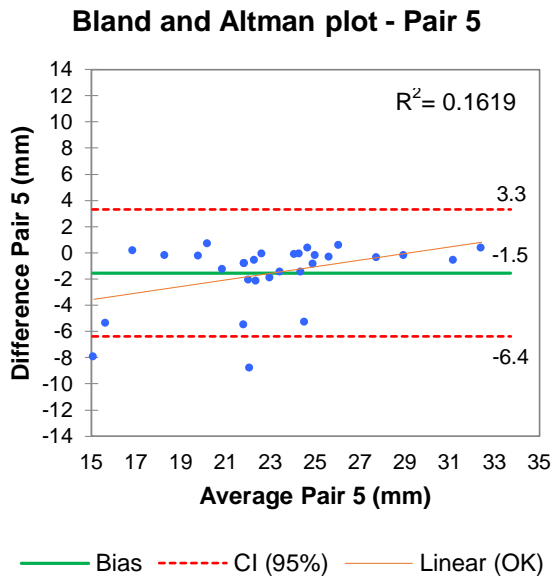


Figure 5. Bland-Altman analysis for biceps muscle thickness between Proc1_R2_M1 and Proc1_R2_M2 (intra-rater measurement reliability).

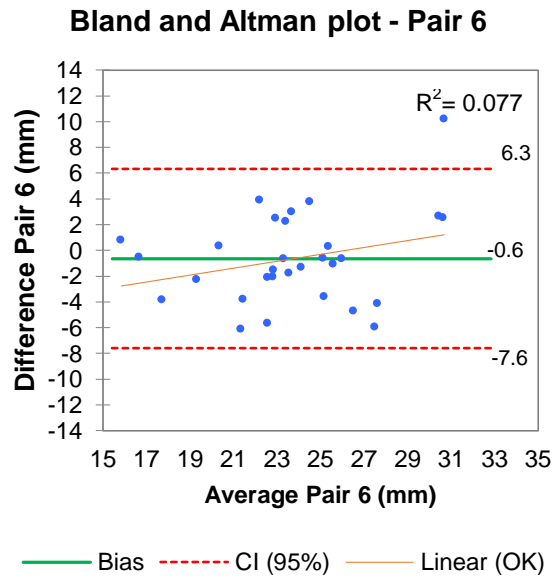


Figure 6. Bland-Altman analysis for biceps muscle thickness between Proc1_R1_M1 and Proc2_R2 (inter-rater procedural reliability).

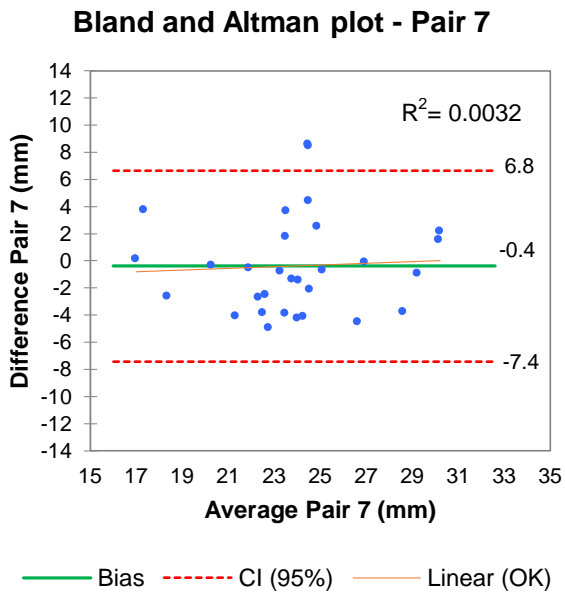


Figure 7. Bland-Altman analysis for biceps muscle thickness between Proc3_R1 and Proc2_R2 (inter-rater procedural reliability).

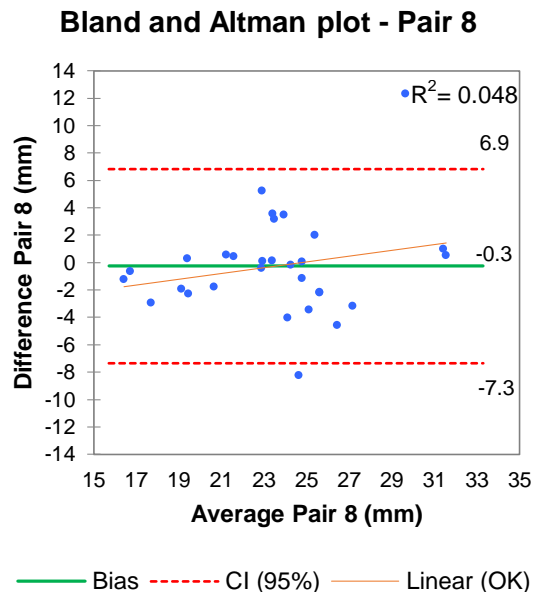


Figure 8. Bland-Altman analysis for biceps muscle thickness between Proc1_R1_M1 and Proc3_R1 (intra-rater procedural reliability).