Original Research

ULTRASOUND

Ultrasound
2019, Vol. 27(4) 233–240

DOI: 10.1177/1742271X19853859
journals.sagepub.com/home/ult
SAGE

Echo intensity reliability between two rectus femoris probe sites

Rodrigo Rabello, Matias Fröhlich, Aline Felicio Bueno, Miriam Allein Zago Marcolino, Thainá De Bona Bernardi, Graciele Sbruzzi and Marco Aurélio Vaz

Abstract

Introduction: The ultrasound technique has been extensively used to measure echo intensity, with the goal of measuring muscle quality, muscle damage, or to detect neuromuscular disorders. However, it is not clear how reliable the technique is when comparing different days, raters, and analysts, or if the reliability is affected by the muscle site where the image is obtained from. The goal of this study was to compare the intra-rater, inter-rater, and inter-analyst reliability of ultrasound measurements obtained from two different sites at the rectus femoris muscle.

Methods: Muscle echo intensity was quantified from ultrasound images acquired at 50% [RF_{50}] and at 70% [RF_{70}] of the thigh length in 32 healthy subjects.

Results: Echo intensity values were higher ($p = 0.0001$) at RF_{50} (61.08 ± 12.04) compared to RF_{70} (57.32 ± 12.58). Reliability was high in both RF_{50} and RF_{70} for all comparisons: intra-rater (ICC = 0.89 and 0.94), inter-rater (ICC = 0.89 and 0.89), and inter-analyst (ICC = 0.98 and 0.99), respectively. However, there were differences ($p < 0.05$) between raters and analysts when obtaining/analyzing echo intensity values in both rectus femoris sites.

Conclusions: The differences in echo intensity values between positions suggest that rectus femoris's structure is not homogeneous, and therefore measurements from different muscle regions should not be used interchangeably. Both sites showed a high reliability, meaning that the measure is accurate if performed by the same experienced rater in different days, if performed by different experienced raters in the same day, and if analyzed by different well-trained analysts, regardless of the evaluated muscle site.

Keywords

Intra-rater, inter-rater, inter-analyst, reproducibility, grayscale analysis

Date received: 7 March 2019; accepted: 8 May 2019

Introduction

The ultrasound (US) technique has been extensively used to measure the muscle quality. By using grayscale analysis, it is possible to obtain a value that represents the amount of white and grey structures in the evaluated muscle, also known as the echo intensity (EI). Structures such as fat and connective tissue are more reflective to the US beam than muscle fibres, resulting in a whiter image the greater their presence. Thus, images with a higher shade of grey suggest a worse muscle quality, while images with a lower shade of grey suggest a better muscle quality.[1,2]

To use a technique with precision, it is fundamental to know if the measurements that it provides are reliable. Reliability evaluation allows us to identify if the values obtained in different situations are different

Escola de Educação Física, Fisioterapia e Dança Laboratório de Pesquisa do Exercício, Universidade Federal do Rio Grande do Sul, Brazil

Corresponding author:
Marco Aurélio Vaz, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil.
Email: marco.vaz@ufrgs.br

because an adaptive process occurred or if they are different because of an embedded error in the technique. Common metrics to quantify reliability are intra-class correlation coefficients (ICC), standard error of measurements (SEM), and minimal difference (MD).[3] In the literature, there are different kinds of reliability. Although the terminology is not constant between authors, three are commonly found. The intra-rater reliability compares the data collected by the same rater in different moments.[4] The inter-rater reliability compares the data collected by different raters in the same day.[5] The inter-analyst reliability uses images regardless of who collected them, comparing the analysis done by different analyzers.[6]

Several studies have evaluated the EI intra-rater reliability, in different muscles (e.g. medial gastrocnemius, quadriceps, hamstrings, and biceps brachii), with different populations (men or women, young or elderly, and healthy or unhealthy) and in different positions (sitting or standing).[4,7–16] However, the results are not consistent amongst them, with reported reliabilities changing considerably from one study to another, with ICC values ranging between 0.31 and 0.96.[4,7–16]

Studies that evaluate inter-rater and inter-analyst reliability are scarce in the literature. Only three studies were found that evaluated EI inter-rater reliability. However, they all evaluated only children, both healthy and diagnosed with Duchenne muscular dystrophy.[5,17,18] As for inter-analyst reliability, to our knowledge, only one study was published, evaluating quadriceps and diaphragm muscles in critical care patients.[6] We were unable to find studies that measured inter-rater and inter-analyst reliability in healthy, young individuals.

One major deficiency of US measurements relates to probe size. Customarily, rectus femoris (RF) US measurements are acquired in the midpoint of the thigh length.[7,12,14,16] However, most probes are not wide enough to capture the whole muscle cross-sectional area in a single transverse image. The extremities are usually cut off, and only part of the muscle can be analyzed. Because EI reliability is a function of the region of interest size,[19] analyzing the whole muscle area is recommended.

Some researchers have found a way to overcome this deficiency, by using the panoramic function that is available in some US systems.[9–11] However, for this panoramic technique to be efficient, the use of an apparatus is necessary to keep the probe moving perpendicular to the skin and in the transverse plane. Furthermore, given the necessary practice and time needed to acquire panoramic US images, Jenkins et al.[8] recommended using single transverse images to quantify EI. Another way to overcome the probe/muscle width issue in the RF muscle is to move the probe distally, where the muscle width is smaller.[20] However, it is not yet clear if EI measurements obtained in a different site of the same muscle are reliable. Therefore, the purpose of this study is to verify, in healthy young subjects, the RF US EI measurements reliability in intra-rater, inter-rater, and inter-analyst comparisons, as well as to compare these reliability measurements when performed in two different RF sites. Furthermore, we wish to determine if these EI values are different between sites and, eventually, if this difference is clinically relevant.

Material and methods

Subjects

Healthy subjects with no injury history on the lower limbs were invited to participate in the study. The number of subjects was determined using the following equation, which indicates the sample size according to the tolerated measurement error for the main analyzed variable

$$n = \frac{Z^2 \times sd^2}{e^2}$$

where n is the sample size, Z is the significance level adopted by the present study (1.96 for $\alpha = 0.05$), sd is the standard deviation of the variable obtained from the literature, and e is the tolerated measurement error (estimated at 5%) applied to the variable mean value obtained from the literature.

The mean and standard deviation values of the right RF muscle EI [48.9 ± 6.9 arbitrary units (a.u.)] were obtained from a study that evaluated a similar population (healthy young subjects) and used a similar methodology.[19] A sample size of 30 individuals was determined as the minimum number of subjects to detect a possible significance. To anticipate eventual sample losses, 32 subjects were recruited for this study. Prior to the study protocol, subjects provided written informed consent. The study was registered and approved by the local ethics committee (CAEE no. 36588914.4.1001.5347, Ethical Approval number: 1.380.427).

Procedures

Subjects visited the laboratory on two occasions, with a one-week interval between them. On the first visit, subjects were evaluated by the first rater (R_1), while on the second visit subjects were re-evaluated by R_1 and were evaluated by the second (R_2) and third (R_3) raters. At the time of the study, R_1 had four years of experience, R_2 had one-year experience, and R_3 had two months experience with the US technique. All raters were instructed to follow the same criteria for probe position and image acquisition, in order that these variables did not influence the results. Raters' order was determined randomly for each subject on the second visit.

Subjects were positioned on a stretcher in the supine position, with the lower limb on top of a steel structure, designed to keep the hip flexed at 60° and the knee flexed at 90° (0° = full extension; Figure 1). Prior to the US measurements, subjects rested for a 20-min period for body fluids stabilization.[21]

A transmission gel was applied to the probe in order to improve acoustic coupling. Three consecutive images were obtained in the transverse plane at each of the two probe sites. In the first site, the probe was placed at 50% of the RF belly length (RF_{50}), and in the second site the probe was placed at 70% of the RF belly length (RF_{70}), both capturing the largest possible muscle cross sectional area (Figure 1).

Ultrasound measurements

All images were obtained using a portable B-mode US device (Vivid-I, General Electric, USA) equipped with a 9 MHz linear-array probe. The depth setting was allowed to be changed at will in order to identify the deep aponeurosis from each participant. However, to guarantee the images comparison obtained at different moments and sites, gain, brightness and contrast were kept at 50% during all evaluations.

All images were analyzed using the Image J software (Version 1.43u, National Institute of Health, Bethesda, MD, USA), by two different analyzers (four years and one-year of experience in US image analysis). The free-hand function was used to select the largest possible muscle area, excluding the aponeurosis. EI was calculated using the mean grey value function, giving a value between 0 (black) and 255 (white).

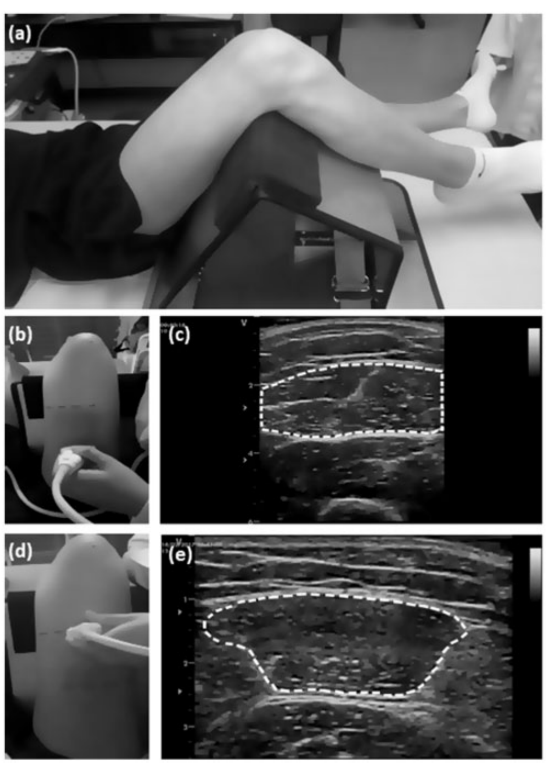

Figure 1. Methodology used for ultrasound evaluation at the two rectus femoris (RF) muscle sites: (a) positioning of the subject for EI evaluation; (b) probe positioned at the RF_{50} site; (c) US image of the muscle on RF_{50} site; (d) probe positioned at the R_{70} site; (e) US image of the muscle on RF_{70} site. Dashed lines indicate how the EI area was determined at the ultrasound images.

Statistical analysis

Intraclass correlation coefficient (ICC) and its 95% confidence interval ($CI_{95\%}$), were calculated using the "2,1" model[3]

$$\mathrm{ICC}\,2,1 = \frac{MS_S - MS_E}{MS_S + (k-1)MS_E + \frac{k(MS_T - MS_E)}{n}}$$

where MS_S is the subjects' mean square, MS_E is the error mean square, MS_T is the mean square total, k is the number of trials, and n is the sample size. Standard error of the measure (SEM) and minimum difference (MD) were calculated to quantify reliability, according to the formulas provided by Weir[3]

$$SEM = SD\sqrt{1} - ICC$$
$$MD = SEM \times 1.96 \times \sqrt{2}$$

Values obtained by the same rater in different days were used to obtain the intra-rater reliability (Figure 2). Values obtained by different raters in the same day were used to obtain the inter-rater reliability (Figure 3). Values obtained by different analysts were used to obtain the inter-analyst reliability. As the Shapiro–Wilk test did not reject the null hypothesis that the data were not normally distributed, parametric tests were used. To verify if the measurements in different sites were different between days, a repeated measures two-way ANOVA was used (factors: moments and sites). To verify if the measurements in different sites were different between raters, a two-way ANOVA was used (factors: raters and sites). A Bonferroni post hoc test was used to identify specific differences. To verify if the measurements in different sites were different between analysts, a two-way ANOVA was used (factors: analysts and sites). To verify the clinical relevance of eventual differences found between measurements, the effect size was calculated adopting the following criteria: <0.2: trivial, >0.2: small;

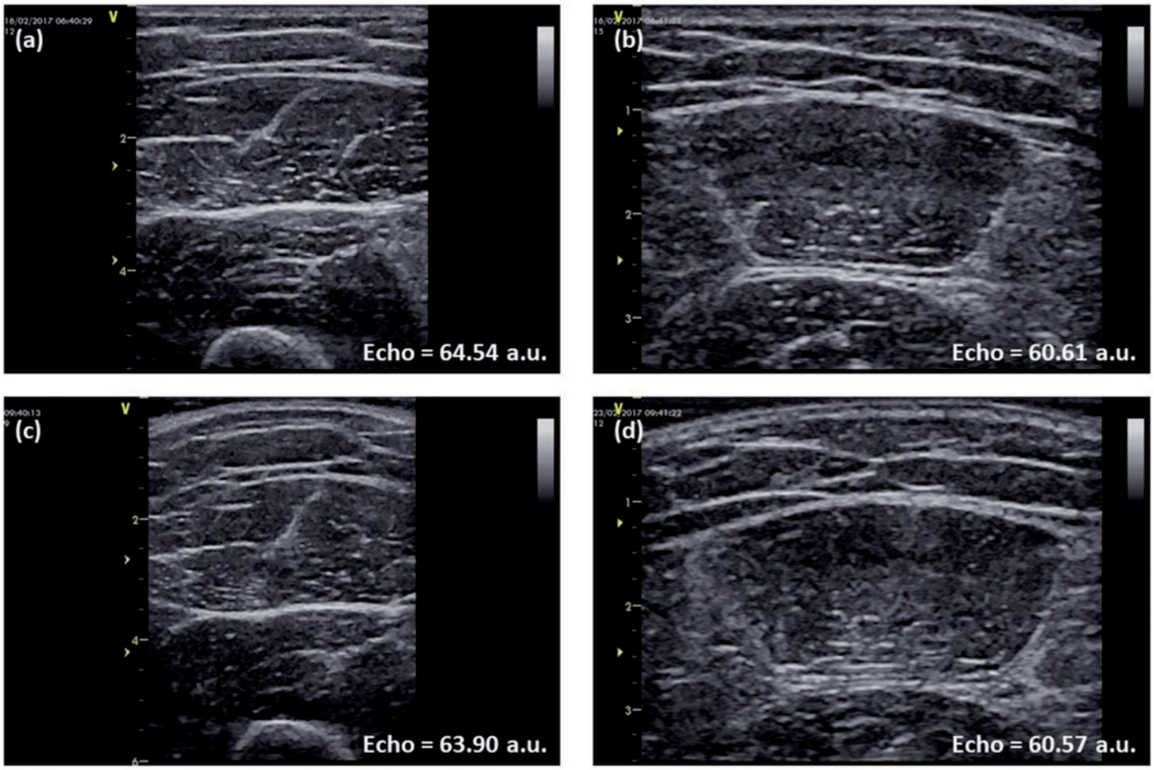

Figure 2. Rectus femoris (RF) ultrasound images from one representative subject, obtained by the same rater (R_1) at two different days and at the two different muscle sites: (a) image obtained by R_1 on the first day on RF_{50} site; (b) image obtained by R_1 on the first day on RF_{70} site; (c) image obtained by R_1 on the second day on RF_{50} site; (d) image obtained by R_1 on the second day on RF_{70} site. Echo intensity values are presented in arbitrary units (a.u.).

>0.50: moderate; > 0.80: large.[22] All analyses were performed with SPSS 20.0 (SPSS Inc., Chicago, IL, USA) software package.

Results

Thirty-two healthy subjects (16 males and 16 females; 26.6 ± 4.9 years) were included. EI mean and standard deviation values are presented in Table 1. For measurements done in different days, RF_{50} and RF_{70} sites were not different ($p = 0.067$). For measurements done by different raters ($p = 0.002$) and analyzed by different analysts ($p < 0.0001$), RF_{50} and RF_{70} sites were different, where the RF_{50} site presented higher echo intensity values. EI values obtained in different days by the same rater were not significantly different ($p = 0.913$). EI values obtained by raters R_1 and R_2 were not significantly different ($p = 1.000$), while values obtained by R_1 and R_3 ($p = 0.006$) and by R_2 and R_3 ($p < 0.0001$) were significantly different, where R_3 obtained smaller values than R_1 and R_2 at both RF_{50} and RF_{70}. Inter-rater EI values analyzed by different analysts showed interaction for analysts and position ($p = 0.028$) and were significantly different for both position ($p = 0.001$; $RF_{50} > RF_{70}$) and analyst ($p < 0.001$). One-way ANOVA for analysts revealed that inter-analysts' comparison was not different for RF_{50} values ($p = 0.472$), but was different for RF_{70} values ($p < 0.001$; $ES = 0.06$), with echo intensity values obtained by Analyst 1 being greater than those obtained by Analyst 2. Effect sizes (calculated through the EI mean and standard deviation values) were classified as small or trivial in all comparisons (0.18–0.41) for both sites.

Reliability statistics are presented in Table 2. For inter-rater comparisons, ICCs were 0.89 and 0.94 for RF_{50} and RF_{70}, respectively, both values indicating a high reliability.[23] SEMs were slightly lower for the RF_{70} site (5.18%) versus for RF_{50} (6.49%), while MDs were slightly higher (18.95% versus 17.94%). For inter-rater comparisons, ICCs were even more similar, 0.89 for RF_{50} and 0.90 for RF_{70}, also indicating high reliability. SEMs and MDs were slightly lower for the RF_{50} site (6.47% and 17.97%) versus RF_{70} (6.84% and 18.95%). For inter-analyst comparisons, ICCs were high for both sites, 0.98 for RF_{50} and 0.99 for RF_{70}. SEMs and MDs were low and slightly smaller for the RF_{70} site (2.14% and 5.92%) compared to RF_{50} (2.52% and 6.98%).

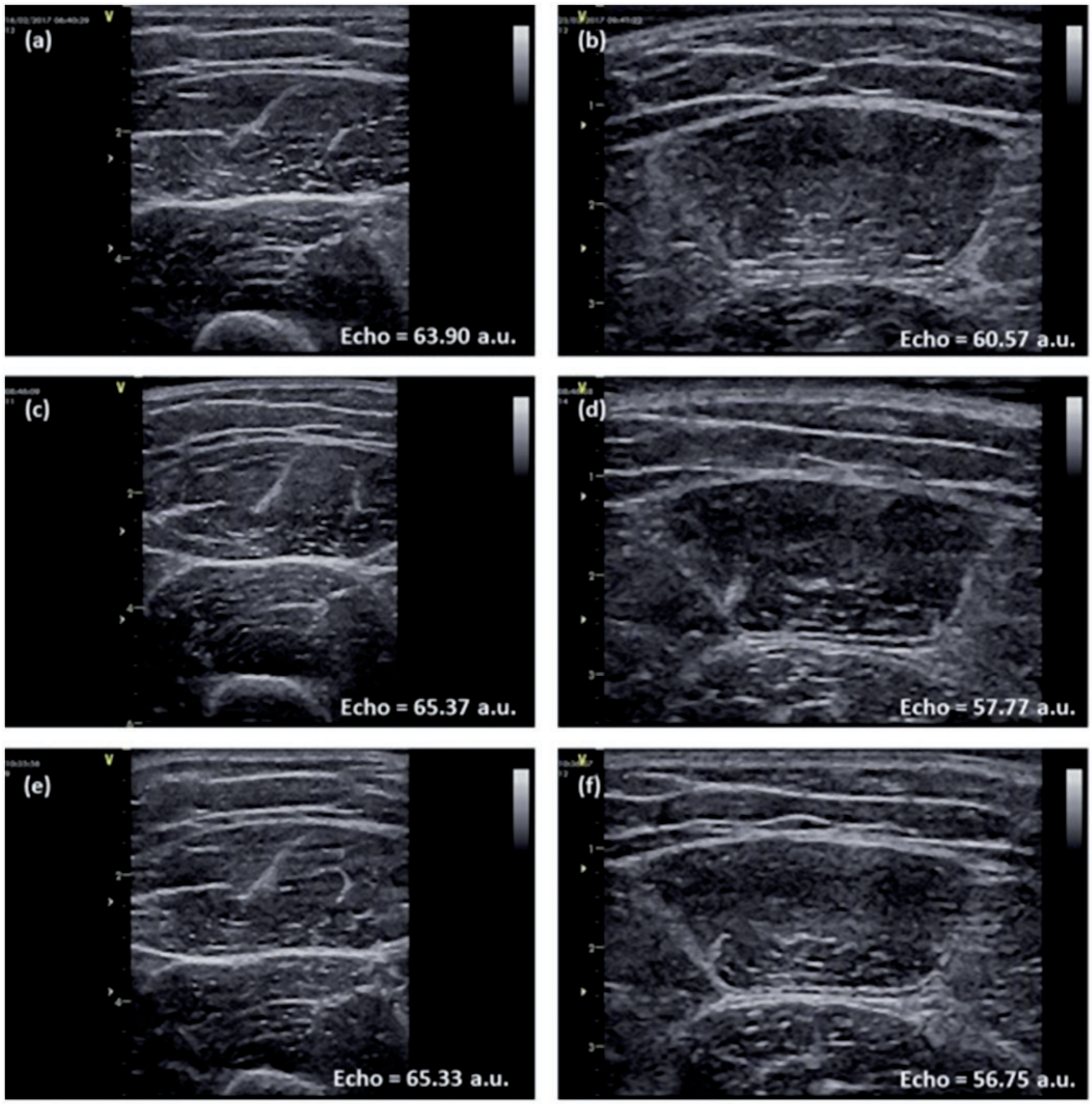

Figure 3. Rectus femoris (RF) ultrasound images from one representative subject, obtained by the three raters at the two different muscle sites: (a) image obtained by R_1 on RF_{50} site; (b) image obtained by R_1 on RF_{70} site; (c) image obtained by R_2 on RF_{50} site; (d) image obtained by R_2 on RF_{70} site; (e) image obtained by R_3 on RF_{50} site; (f) image obtained by R_3 on RF_{70} site. Echo intensity values are presented in arbitrary units (a.u.).

Discussion

Because of its low cost, high effectiveness, and accessibility, US has been used in the majority of studies and in clinical settings to assess EI. However, US is a technique that has been said to be rater dependent.[24] If so, then it could not be considered a reliable technique. The main purpose of the present study was to quantify the reliability and the US technique error of measurement in intra-rater, inter-rater, and inter-analyst comparisons. Our results showed that US images can be taken with a very high reliability in different moments by the same rater, by different raters and can be analyzed by different analysts. However, this needs to be done with caution.

Between the aforementioned comparisons, intra-rater is the most commonly found, with US measurements taken from different populations and different muscles. Specifically, in the RF, ICCs ranged between 0.31 and 0.91.[7,9,12–14,16] Ruas et al.[12] reported SEMs of 8.81% (MDs = 24.44%) of the mean EI value, Tomko et al.[16] found SEM values ranging from 6.46% to 8.12% of the mean EI, while Santos and Armada-da-Silva[13] reported SEMs ranging from 2.06 to 3.61 (a.u.).

In this study, for intra-rater comparisons, the ICCs were 0.89 and 0.94 for the RF_{50} and RF_{70}, respectively,

Table 1. EI values for the analysis made in different days, by different raters and by different analysts

	RF_{50}	RF_{70}	Position (p-value)	Position (effect size)	
Day 1	60.80 ± 12.56	57.72 ± 13.18	0.067	0.24	Small
Day 2	61.59 ± 11.26	58.28 ± 13.32	0.067	0.27	Small
R_1	61.09 ± 12.37[a]	58.84 ± 13.19[a]	0.002	0.18	Trivial
R_2	62.20 ± 12.27[b]	57.32 ± 11.31[b]	0.002	0.41	Small
R_3	59.71 ± 11.99	55.23 ± 11.74	0.002	0.38	Small
Analyst 1	61.16 ± 12.05	57.72 ± 12.65	<0.0001	0.28	Small
Analyst 2	61.00 ± 12.03[c]	56.93 ± 12.52[c]	<0.0001	0.33	Small

Values are means ± standard deviations. There was no significant difference between days ($p = 0.913$) and no significant difference between R_1 and R_2 ($p > 0.05$) in both sites.
[a]Significant difference between R_1 and R_3 ($p = 0.006$).
[b]Significant difference between R_2 and R_3 ($p < 0.0001$).
[c]Significant difference between analysts ($p = 0.001$).
Days and rater values are a mean of both analysts' results. Analysts' values are means of all days and raters analyzed by them.

Table 2. Intra-rater, inter-rater, and inter-analyst reliability statistics for EI

	RF_{50}			RF_{70}		
	Intra-rater	Inter-rater	Inter-analyst	Intra-rater	Inter-rater	Inter-analyst
ICC	0.89	0.89	0.98	0.94	0.90	0.99
$CI_{95\%}$	0.78–0.94	0.82–0.94	0.97–0.99	0.89–0.97	0.82–0.94	0.98–0.99
SEM (AU)	3.97	3.95	1.54	3.01	3.87	1.21
SEM (%)	6.49	6.47	2.52	5.18	6.84	2.14
MD (AU)	10.98	10.96	4.26	8.37	10.75	3.36
MD (%)	17.94	17.97	6.98	14.37	18.95	5.92

ICC: intra-class coefficient correlation; 95% CI: 95% confidence interval for ICC; SEM (a.u.): standard error of measurement expressed in arbitrary units; SEM (%): standard error of measurement expressed as a percentage of the grand mean; MD (a.u.): minimum difference needed to be considered real expressed in arbitrary units; MD (%): minimum difference needed to be considered real expressed as a percentage of the grand mean.

a reliability similar to the highest results found previously in the literature. The SEM values of our study were slightly lower than the ones previously reported by Ruas et al.[12] and Santos and Armada-da-Silva[13] and within the range reported by Tomko et al.,[16] while the MDs were even lower.[12] These results indicate that RF EI measurements are reliable when obtained in different days by the same rater, meaning that, if results are different between sessions, they are likely because of a real change in the muscle structure.

In the present study, the inter-rater ICCs were 0.89 and 0.90, while the SEMs were 6.47% and 6.84%, and the MDs were 17.97% and 18.95% of the mean for RF_{50} and RF_{70}, respectively. These numbers show high inter-rater reliability and agree with previous studies conducted with children that obtained ICCs from 0.82 and 0.99.[5,17,18] However, none of the previous studies reported SEMs or MDs. The relevance of a high inter-rater reliability can be exemplified by a scenario where a subject can be evaluated by different raters or clinicians, not depending on the times when the rater is available, and allowing a greater ease in conducting this evaluation in clinical settings.

Inter-analyst ICCs were almost perfect, 0.98 for RF_{50} and 0.99 for RF_{70}. SEMs were 2.52% and 2.14%, while MDs were 6.98% and 5.92% for RF_{50} and RF_{70}, respectively. These results indicate that there is great reliability of EI images analyzed by different analysts, suggesting that analysis can be safely done by different analyzers. Sarwal et al.[6] found similar values for quadriceps and abdominal muscles, finding ICC values ranging from 0.84 to 0.99 while evaluating a critically ill population.

In between-days comparisons, EI values were not different neither in different days nor in different sites, suggesting that the absolute EI value is not influenced by being obtained in different days or sites, when acquired by the same rater. Between raters and between analysts, EI values were different when comparing the two sites, suggesting that the difference in values between RF_{50} and RF_{70} are influenced by raters and analysts.

By identifying specific differences between EI values obtained by the different raters, it can be observed that measurements performed by raters R_1 and R_2 were not different, while the measurements were different when comparing R_3 with both R_1 and R_2. These results could be partially explained by the rater's practice time of the technique, as R_3 had only two months of practice prior to data collection, while R_1 had four years and R_2 had one year of practice. This suggests that, although inter-rater reliability is high, the experience with the technique of the different raters could influence the obtained values.

EI values obtained by different analysts were significantly different, suggesting that they are influenced by whoever analyzes the images, regardless of having received the exact same instruction on how to analyze the images. Effect sizes obtained when comparing both sites were considered small for all days, raters, and analysts. Overall, we can observe that EI values obtained by different raters and different analysts can be influenced by the muscle site where the image is obtained. Thus, one should have caution in switching raters/analysts. However, if switching raters/analysts is necessary as usually observed in clinical practice, additional analysis of individual changes in light of the MDs may be warranted.

The reliability between the two sites was similar, which suggests that both can be used to evaluate EI in RF. The RF_{70} site usually allows for the totality of the RF area to be visualized, and it does not require additional practice or apparatus.[8] Therefore, measurements in the RF_{70} site allow for faster evaluations, while still being reliable. Thus, researchers and clinicians might consider adopting a 70% thigh length probe positioning when EI is the measured variable. Nevertheless, further studies evaluating the effect of muscle adaptation due to training or disuse on different RF sites should be conducted in order to verify if the level of change in EI is also similar. A between-sites difference in longitudinal studies might show heterogeneity between different muscle sites adaptation, which might also have clinical and functional implications during rehabilitation programs.

Possible limitations of the study include the difference in experience of the three different raters, which apparently did influence their measurements, and therefore longer training periods in the use of US are suggested. Similarly, despite the small differences between the two analysts' outcome values, additional training in the US image analysis might be necessary. Another possible limitation regards the fact that our participants were young healthy subjects, whereas in clinical settings patients will often display a greater variability in muscle quality, a factor that might reduce the reliability of these measures, and which reinforces the need of a good training of raters and analysts prior to data collection/analysis.

Conclusions

RF EI measurements are accurate if performed by the same examiner in different days or if performed by different experienced examiners in the same day. It is also accurate if different analysts perform the image analysis. However, nonexperienced raters need to undergo a longer training period before their EI images can be considered reliable compared to more experienced raters. Similarly, analysts need to undergo US image analysis training in order to reduce the minimal difference between analysts. RF_{50} and RF_{70} sites showed a high and comparable reliability, meaning that the measurement can be done safely in both sites. However, the differences in EI values between positions suggest that EI is not homogeneous in RF, indicating that measurements from different muscle regions should not be used interchangeably to measure EI. Despite the small between-sites differences, RF_{70} might be better compared to RF_{50} as it allows for the inclusion of a larger muscle area to determine EI in most subjects.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by CNPq-Brazil (Grant number 458838/2013-6).

Ethics approval

Comitê de Ética em Pesquisa da UFRGS
Av. Paulo Gama
Porto Alegre/RS - CEP: 90040-060
Ethical approval number: 1.380.427

Guarantor

MAV

Contributors

MF, GS, and MAV conceived and designed the research. RR, MF, AFB, MAZM, and TBB conducted the experiments. RR and MF analyzed the data. RR and MF wrote the article. All authors read, reviewed, and approved the article.

ORCID iD

Rodrigo Rabello https://orcid.org/0000-0003-0665-1835

References

1. Walker FO, Cartwright MS, Wiesler ER, et al. Ultrasound of nerve and muscle. *Clin Neurophysiol* 2004; 115: 495–507.
2. Pillen S, Arts IM and Zwarts MJ. Muscle ultrasound in neuromuscular disorders. *Muscle Nerve* 2008; 37: 679–693.
3. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res* 2005; 19: 231–240.
4. Vieira A, Siqueira AF, Ferreira-Junior JB, et al. Ultrasound imaging in women's arm flexor muscles: intra-rater reliability of muscle thickness and echo intensity. *Braz J Phys Ther* 2016; 20: 535–542.
5. Zaidman CM, Wu JS, Wilder S, et al. Minimal training is required to reliably perform quantitative ultrasound of muscle. *Muscle Nerve* 2014; 50: 124–128.
6. Sarwal A, Parry SM, Berry MJ, et al. Interobserver reliability of quantitative muscle sonographic analysis in the critically ill population. *J Ultrasound Med* 2015; 34: 1191–1200.
7. Fukumoto Y, Ikezoe T, Yamada Y, et al. Skeletal muscle quality assessed from echo intensity is associated with muscle strength of middle-aged and elderly persons. *Eur J Appl Physiol* 2012; 112: 1519–1525.
8. Jenkins ND, Miller JM, Buckner SL, et al. Test-retest reliability of single transverse versus panoramic ultrasound imaging for muscle size and echo intensity of the biceps brachii. *Ultrasound Med Biol* 2015; 41: 1584–1591.
9. Melvin MN, Smith-Ryan AE, Wingfield HL, et al. Evaluation of muscle quality reliability and racial differences in body composition of overweight individuals. *Ultrasound Med Biol* 2014; 40: 1973–1979.
10. Palmer TB, Akehi K, Thiele RM, et al. Reliability of panoramic ultrasound imaging in simultaneously examining muscle size and quality of the hamstring muscles in young, healthy males and females. *Ultrasound Med Biol* 2015; 41: 675–684.
11. Rosenberg JG, Ryan ED, Sobolewski EJ, et al. Reliability of panoramic ultrasound imaging to simultaneously examine muscle size and quality of the medial gastrocnemius. *Muscle Nerve* 2014; 49: 736–740.
12. Ruas C, Pinto R, Lima C, et al. Test-retest reliability of muscle thickness, echo-intensity and cross sectional area of quadriceps and hamstrings muscle groups using B-mode ultrasound. *Intern J Kinesiol and Sports Sci* 2017; 5: 6.
13. Santos R and Armada-da-Silva PAS. Reproducibility of ultrasound-derived muscle thickness and echo-intensity for the entire quadriceps femoris muscle. *Radiography (Lond)* 2017; 23: e51–e61.
14. Strasser EM, Draskovits T, Praschak M, et al. Association between ultrasound measurements of muscle thickness, pennation angle, echogenicity and skeletal muscle strength in the elderly. *Age (Dordr)* 2013; 35: 2377–2388.
15. Tanaka NI, Ogawa M, Yoshiko A, et al. Reliability of size and echo intensity of abdominal skeletal muscles using extended field-of-view ultrasound imaging. *Eur J Appl Physiol* 2017; 117: 2263–2270.
16. Tomko PM, Muddle TW, Magrini MA, et al. Reliability and differences in quadriceps femoris muscle morphology using ultrasonography: the effects of body position and rest time. *Ultrasound* 2018; 26: 214–221.
17. Pigula-Tresansky AJ, Wu JS, Kapur K, et al. Muscle compression improves reliability of ultrasound echo intensity. *Muscle Nerve* 2018; 57: 423–429.
18. Shklyar I, Geisbush TR, Mijialovic AS, et al. Quantitative muscle ultrasound in Duchenne muscular dystrophy: a comparison of techniques. *Muscle Nerve* 2015; 51: 207–213.
19. Caresio C, Molinari F, Emanuel G, et al. Muscle echo intensity: reliability and conditioning factors. *Clin Physiol Funct Imaging* 2015; 35: 393–403.
20. e Lima KM, da Matta TT and de Oliveira LF. Reliability of the rectus femoris muscle cross-sectional area measurements by ultrasonography. *Clin Physiol Funct Imaging* 2012; 32: 221–226.
21. Berg HE, Tedner B and Tesch PA. Changes in lower limb muscle cross-sectional area and tissue fluid volume after transition from standing to supine. *Acta Physiol Scand* 1993; 148: 379–385.
22. Cohen J. *Statistical power analysis for the behavioral sciences*, 2nd ed. Hillsdale, NJ: Lawrence Earlbaum Associates, 1988.
23. Katz JN, Larson MG, Phillips CB, et al. Comparative measurement sensitivity of short and longer health status instruments. *Med Care* 1992; 30: 917–925.
24. Harris-Love MO, Monfaredi R, Ismail C, et al. Quantitative ultrasound: measurement considerations for the assessment of muscular dystrophy and sarcopenia. *Front Aging Neurosci* 2014; 6: 172.