Inter- and intra-rater reliability of ultrasound-derived muscle quantification

Christopher Cleary
University of Kansas
Omid Nabavizadeh
University of Kansas
Kaycie Young
University of Kansas
Ashley Herda (a.herda@ku.edu)
University of Kansas

Research Article

Keywords: Intra-class correlation coefficient, muscle thickness, cross-sectional area, echo intensity

Posted Date: May 6th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-465652/v1

Abstract

Ultrasound devices are now commonplace in sports medicine facilities due to their ease of use and validity to assess skeletal muscle characteristics. The aim of the current study was to assess the reliability of ultrasound skeletal muscle image analysis across multiple raters. Vastus lateralis (VL), rectus femoris (RF), and first dorsal interosseus (FDI) images were analyzed in ImageJ by three raters to determine muscle thickness (MT), cross-sectional area (CSA), and echo-intensity (EI). Intra-class correlation coefficients (ICC) between (inter-rater) and within (intra-rater) raters, the standard error of the measurement (SEM) and minimal difference needed to be considered real (MD) were calculated. Inter-rater reliability was high for the selected outcomes of the VL and RF (ICC: 0.987–0.999), while the FDI was lower (0.614–0.962). The MD was small in each muscle and outcome. Intra-rater reliability ranged from 0.961–0.999 for all three raters of differing experience with ImageJ. Providing specific directions to analyze ultrasound images results in reliable ultrasound measurements for the FDI, RF, and VL. The FDI may have had lower ICCs due to the uneven nature and borders. Future studies should continue to investigate the reliability of ultrasound measurements across more raters, various skeletal muscles, and independently per lab group.

Introduction

Ultrasound imaging is a non-invasive and widely available modality to assess skeletal muscle characteristics such as muscle thickness (MT), muscle cross-sectional area (CSA), and echo intensity (EI) [1, 2]. Ultrasound devices are less cumbersome, inexpensive, portable, and potentially safer than computed topography (CT) or magnetic resonance imaging (MRI) when used for similar applications. For skeletal muscle assessments, ultrasound imaging is useful to evaluate the efficacy of a rehabilitation program following surgery [3], monitor rates of atrophy during bed rest [4], or assess adaptations derived from resistance training programs [5, 6]. Traditional ultrasound devices have a small field of view which could potentially limit the ability to adequately measure skeletal muscle CSA in larger muscles. However, advanced scanning technology that allows an extended field of view, deemed panoramic ultrasound imaging, has been recently utilized to allow muscles with larger CSAs to be examined [7, 8].

Panoramic ultrasound has also been previously validated to other skeletal muscle imaging modalities, such as MRI and CT, demonstrating comparable results [9]. Specifically, Ahtiainen et al. examined the vastus lateralis (VL) muscle CSA via MRI and panoramic ultrasound before and after a 21-week lower-body resistance training intervention and was determined to be valid and reliable to measure changes in muscle CSA when compared to MRI [5]. Additionally, Noorkoiv et al. compared panoramic ultrasound to CT for the quadriceps muscle CSA and observed minimal (1.3-2.5%) differences between the two techniques, with excellent intra-class correlation coefficients (ICC) of the two methods ranging from 0.951 to 0.998 [10]. Similar results were reported by Scott et al. in which ultrasound had high levels of agreement for muscle CSA with MRI following 70 days of bed rest [8]. The agreement among these differing methods suggests that any method may be used as they are available and fit the needs of the specific diagnostic image goals and levels of sensitivity from each device. It also allows for smallerscaled laboratories to cost-effectively quantify muscle in research settings when access to MRI or CT devices is lacking [11].

Despite the increasing prevalence of ultrasound devices, it is vital to continually investigate the reliability measurements obtained across and within raters [4, 12]. Prior to conducting ICC analyses, it has been proposed that a repeated-measures analysis of variance (ANOVA) be conducted first [13]. This indicates if systematic variability is present and if mean differences between trials or raters are significantly different. However, Weir has suggested that just although the ANOVA may indicate significant differences, it does not necessarily mean that the ICCs are unreliable, as small changes in the mean differences may result in a significant ANOVA [13]. Previous studies have reported high test-retest reliability across days for panoramic ultrasound image acquisition [1]. For the reliability between individual raters (inter-rater reliability), ultrasound measurements have high reliability, as the quadriceps MT has been shown to have an ICC of 0.98 [4], CSA has an ICC of 0.987 for the VL and 0.963 for the rectus femoris (RF) between multiple raters [2]. The investigator that conducts the scan must be accurate and consistent as suggested by Teyhen et al. [14], as well as the individual(s) conducting the image analysis. Further, the reliability of panoramic ultrasound for various skeletal muscles of the lower and upper body have not been investigated collectively. For example, the first dorsal interosseus (FDI) has commonly been assessed to investigate neuromuscular function in diseased and healthy populations [15, 16], yet no studies have assessed the reliability of FDI panoramic ultrasound, nor compare the ICCs of upper and lower limb muscles. The lower limb muscles are large and have well-defined borders, as discussed by Pillen, yet some thoracic and upper body muscles are not as easy to differentiate due to their size and arrangement [17]. These differences may lead to differing reliability of image analysis.

Furthermore, if specific and established guidelines and instructions on how to analyze a panoramic ultrasound image are provided to raters, then the results of multiple studies can be easily compared. Thus, it is imperative to ensure these individuals produce consistent measurements if there are multiple analysts or raters working on a large-scale research study. Previous studies have suggested that only two raters are not enough to quantify accurate ultrasound mean values [18, 19] and that up to three raters may result in more accurate analyses. It is imperative to assess the reliability of ultrasound analyses as laboratory groups may bring in new students or laboratory assistants over time and to increase the accuracy of these new raters, specified instructions may be utilized. Therefore, the purpose of this study was to evaluate the intra- and inter-rater reliability of ultrasound image analysis from the vastus lateralis, rectus femoris, and first dorsal interosseous captured by a single investigator and analyzed using pre-determined methodology among multiple raters.

Methods

This study was a retrospective design with the purpose of analyzing previously collected ultrasound images of the VL, RF, and FDI and comparing these analyses across three raters. Each rater analyzed the images in triplicate for MT, CSA, and EI. Images were acquired by multiple investigators using the same image acquisition methods. Panoramic ultrasound images of the VL (n = 15), RF (n = 10), and the FDI (n

= 12) were assessed. All images were obtained from previous projects conducted in our laboratory (unpublished data) [20]. Descriptive characteristics of the participants were not obtained for the purposes of this study. All images were previously acquired from participants that signed written statements of informed consent from unrelated studies, all studies and procedures were approved by the University of Kansas Human Research Protection Program (HRPP) in accordance with the Declaration of Helsinki. All experiments were performed in accordance with relevant guidelines and regulations (#STUDY000146569).

Procedures

Equipment and settings. For the VL and RF, the images were captured using a Logiq e ultrasound device in B-mode (GE Healthcare, Wauwatosa, WI, USA) with a linear array transducer (Model L4-12t-RS, 4.2-30 MHz). Images were acquired on the right leg for all participants while the subjects laid supine on the examination table. The subjects rested in the supine position for 15 minutes prior to recording the images to allow for the shifting of fluids. The scan depth was set to 6 cm, gain was 68 dB, and transducer frequency was 10 MHz to optimize image quality and was held constant across all subjects. The RF images were captured at 50% of the distance between the greater trochanter and the superior pole of the patella [21] while VL images were captured at 50% of the length between the greater trochanter of the femur and lateral epicondyle of the femur [5]. For the FDI, the images were captured using a Logiq e ultrasound device in B-mode (GE Healthcare UK, Chalfont, UK) with a linear array transducer (Model 12L-RS, 5-13 MHz). All images were assessed on the FDI of the right hand. Subjects were seated during all measurements and the right hand was placed on the examination table on top of foam pads to elevate the hand from the table. Scan depth was set at 4 cm, gain was 38 dB and transducer frequency was 12 MHz. Images were captured at the midpoint of the origin and insertion of the FDI as identified with a longitudinal scan [20]. The probe was positioned perpendicularly to the second metacarpal during image acquisition. Prior to each ultrasound measurement, a copious amount of water-soluble transmission gel was applied to the skin and the transducer to enhance image quality and provide acoustic coupling. The transducer was moved manually across the selected muscle in the transverse plane with a constant, minimal pressure. All images were saved to an external storage device and exported for subsequent analyses.

Image analysis. Three raters manually traced the VL, RF, and FDI for MT, CSA, and EI. These analyses were conducted utilizing readily available software from the National Institutes of Health (NIH), ImageJ (NIH, Bethesda, MD, USA). Each investigator traced the images three times with intra-rater reliability performed separately for each rater on MT, CSA, and EI of the three muscles. The average of the three scans was used to calculate inter-rater reliability for MT, CSA, and EI of the three skeletal muscles. Prior to analysis, each image was scaled from pixels to cm with ImageJ’s straight-line function using a known distance of 2 cm. MT was determined as the greatest thickness at the center-point of the muscle belly, perpendicular to the femur for the VL and RF and perpendicular to the second metacarpal for the FDI, using ImageJ’s straight-line function. Muscle CSA was assessed using ImageJ’s polygon function by manually tracing the border of the muscle belly. EI was determined as the mean gray-scale analysis in

ImageJ of the muscle's selected region of interest (AU) (black = 0, white = 255). Care was taken to ensure that muscle fascia was not included in the MT and CSA analyses. Figure 1 (a, b, and c) depicts the tracing of each of the muscles; RF, VL, and FDI, respectively.

Statistical Analyses

Repeated-measures ANOVAs were first conducted to examine for systematic variability in CSA, MT, and EI for each muscle [13]. Intra-class correlation coefficients (ICC), the standard error of measurement (SEM), and the minimal difference (MD) needed to be considered real were calculated for each muscle and variable as well [1, 13]. The ICC model "2,*k*" was utilized to calculate the ICCs for both intra- and inter-rater reliability [13, 22]. Model 2,*k* was chosen as it has been suggested that the ICCs generated with this model can be generalized to other laboratories and testers and uses greater than two raters and time points [1]. Inter- and intra-rater reliability was performed for the CSA, MT, and EI of the VL, RF, and FDI. The SEM was calculated from the square-root of the mean square error (MS_E). The MD needed to be considered real was calculated from the SEM via equations put forth by Weir [13]. The calculations for ICC, SEM, and MD were performed using a custom-written Microsoft Excel spreadsheet (Microsoft Excel, Microsoft, Redmond, WA, USA). The ANOVA was conducted in IBM SPSS v 27.0 (IBM, Armonk, NY, USA). An alpha level of $p < 0.05$ was considered statistically significant for all analyses.

Results

Inter-rater reliability statistics measured by $ICC_{2,k}$, SEM, and MD for the MT, CSA, and EI of all three muscles are presented in Table 1. Systematic variability was present in MT of the RF and FDI ($p = 0.031$ and $p = 0.043$, respectively), CSA of the VL ($p = 0.006$), and EI of the VL and RF ($p < 0.001$ and $p = 0.006$, respectively). Systematic variability was not present in any of the other outcomes for the selected muscles ($p = 0.064 – 0.547$). Inter-rater reliability was found to be highly consistent in the VL ($ICC_{2,k}$ range = 0.993 – 0.998) and the RF ($ICC_{2,k}$ range = 0.988 – 0.999) across the three raters. For the FDI, MT had an $ICC_{2,k}$ of 0.614, while CSA and EI were 0.842 and 0.962, respectively.

Table 1. Inter-rater reliability for muscle thickness, cross-sectional area, and echo intensity.

		VL (n = 15)	RF (n = 10)	FDI (n = 12)
MT	*p*	0.497	0.031*	0.043*
	ICC $_{2,k}$	0.993	0.987	0.614
	SEM (cm)	0.048	0.067	0.075
	MD (cm)	0.133	0.187	0.208
CSA	*p*	0.006*	0.064	0.387
	ICC $_{2,k}$	0.987	0.997	0.842
	SEM (cm^2)	0.796	0.288	0.223
	MD (cm^2)	2.206	0.799	0.619
EI	*p*	0.001*	0.006*	0.547
	ICC $_{2,k}$	0.997	0.999	0.962
	SEM (AU)	0.937	0.603	1.751
	MD (AU)	2.596	1.670	4.854

EI = echo intensity; MT = muscle thickness; CSA = cross-sectional area; SEM = standard error of the measurement; MD = minimal difference needed to be considered real; VL = vastus lateralis; RF = rectus femoris; FDI = first dorsal interosseous; AU = arbitrary units.

* = significant systematic variability from the ANOVA between the three raters ($p < 0.05$).

Table 2. Depiction of each rater's average measurements for each outcome and muscle.

	VL			RF			FDI		
	MT (cm)	CSA (cm^2)	EI (AU)	MT (cm)	CSA (cm^2)	EI (AU)	MT (cm)	CSA (cm^2)	EI (AU)
Rater 1	2.30 ± 0.33	26.67 ± 5.79	61.52 ± 17.50	2.23 ± 0.48	10.57 ± 3.78	69.65 ± 23.40	0.88 ± 0.09	1.93 ± 0.35	48.55 ± 4.77
Rater 2	2.31 ± 0.32	27.11 ± 5.39	62.10 ± 17.89	2.31 ± 0.48	10.89 ± 3.84	69.19 ± 23.52	0.95 ± 0.13	1.82 ± 0.41	47.80 ± 5.66
Rater 3	2.29 ± 0.33	26.08 ± 5.73	60.57 ± 17.51	2.23 ± 0.47	10.78 ± 3.69	68.66 ± 23.27	0.88 ± 0.08	1.94 ± 0.32	48.41 ± 4.68

VL = vastus lateralis; RF = rectus femoris; FDI = first dorsal interosseus; MT = muscle thickness; CSA = cross-sectional area; EI = echo-intensity.

Table 2 depicts the raters' average of three measurements for each outcome and muscle. Within each rater, intra-rater reliability ranged from an $ICC_{2,k}$ of 0.961 – 0.999. The ANOVA indicated significant differences were observed for Rater 1 in VL EI ($p = 0.002$), Rater 2 in MT and CSA of the VL ($p = 0.016$ and $p = 0.010$, respectively), and Rater 3 in VL CSA and RF EI ($p = 0.019$ and $p = 0.032$, respectively). Systematic variability was not present in any of the other outcomes for any of the three raters ($p > 0.05$). The SEM values for MT ranged from 0.048 cm to 0.075 cm and the MD ranged from 0.133 cm to 0.208 cm. For CSA, SEM ranged from 0.223 cm^2 to 0.796 cm^2, while the MD ranged from 0.619 cm^2 to 2.206 cm^2. Lastly, EI had an SEM ranging from 0.603 arbitrary units (AU) to 1.751 AU and MD AU of 1.670 to 4.854 AU.

Table 3 presents the SEM and expressed as a percent of the mean. Table 4 reports the MD data for each outcome and muscle for each rater as well as the average MD across raters.

Table 3. Standard error of the measurement (SEM) values for each rater and expressed as a percentage of the mean in parentheses.

	VL			RF			FDI		
	MT (cm)	**CSA (cm^2)**	**EI (AU)**	**MT (cm)**	**CSA (cm^2)**	**EI (AU)**	**MT (cm)**	**CSA (cm^2)**	**EI (AU)**
Rater 1	0.075 (3.3%)	0.888 (3.3%)	0.826 (1.3%)	0.056 (2.5%)	0.554 (5.2%)	0.622 (0.9%)	0.024 (2.7%)	0.111 (5.8%)	0.803 (1.6%)
Rater 2	0.055 (2.3%)	0.966 (3.6%)	1.41 (2.3%)	0.073 (3.2%)	0.259 (2.4%)	0.416 (0.6%)	0.039 (4.1%)	0.02 (1.1%)	0.292 (0.6%)
Rater 3	0.037 (1.6%)	0.188 (0.7%)	0.389 (0.6%)	0.029 (1.3%)	0.121 (1.1%)	0.226 (0.3%)	0.028 (3.2%)	0.026 (1.3%)	0.995 (2.1%)

MT = muscle thickness; CSA = cross-sectional area; EI = echo intensity; VL = vastus lateralis; RF = rectus femoris; FDI = first dorsal interosseus; AU = arbitrary units.

Table 4. Minimal difference needed to be considered real (MD) for each rater and the mean ± SD.

	VL			RF			FDI		
	MT (cm)	CSA (cm^2)	EI (AU)	MT (cm)	CSA (cm^2)	EI (AU)	MT (cm)	CSA (cm^2)	EI (AU)
Rater 1	0.208	2.462	2.290	0.156	1.537	1.725	0.067	0.307	2.226
Rater 2	0.154	2.676	3.908	0.203	0.717	1.154	0.109	0.055	0.808
Rater 3	0.102	0.520	1.078	0.080	0.336	0.738	0.079	0.072	2.758
Mean MD	0.155 ± 0.053	1.886 ± 1.188	2.425 ± 1.420	0.146 ± 0.062	0.863 ± 0.614	1.206 ± 0.496	0.085 ± 0.022	0.145 ± 0.141	1.931 ± 1.008

MT = muscle thickness; CSA = cross-sectional area; EI = echo intensity; VL = vastus lateralis; RF = rectus femoris; FDI = first dorsal interosseus; AU = arbitrary units.

Discussion

The primary outcomes of this study included high reliability of image analysis between multiple raters. The purpose of this study was to assess the inter-rater and intra-rater reliability of three raters assessing MT, CSA, and EI of the VL, RF, and FDI. Despite the high reliability, measured by $ICC_{2,k}$, significant systematic variability between the three raters was present in RF and FDI MT measurements, muscle CSA of the VL, and EI of the VL and RF. The FDI had lower ICC values across the three raters than the other two muscles. Although the FDI had the smallest MT and muscle CSA, it had the largest SEM for MT of 0.075 cm. Each of the three raters had high ICC values for each outcome and muscle, and low SEM and MD values.

Although systematic variability between the raters was present in some of the outcomes for the selected muscles, the inter-rater ICCs were still high. Weir suggested that for an ANOVA to have significant differences, the mean differences between trials (or raters in the present study) must be large, while the error term is small, or both [13]. In the present study, the high ICCs and small error terms suggest that raters were very consistent in their assessment of each image. Similar to Filippo et al., the variability within a rater in the present study is less than between raters [4]. Further, Reis et al. reported the use of multiple raters is acceptable because the reliability is within predefined ICC limits of 0.75-0.90 as good or >0.90 as excellent [23]. Although the interrater ICC values were 'good to excellent’ for most variables (0.614 – 0.999), it is also useful to include information on the SEM and the MD when assessing reliability, as the SEM is an absolute index that is independent from the population it was derived from, unlike the ICC [24]. Additionally, the MD is calculated from the SEM and reflects the difference in a score needed to be considered meaningful beyond the normal measurement error [13].

The SEM values in the present study for two quadriceps femoris muscles, RF and VL, ranged from 0.048 – 0.067 cm for MT, 0.288 – 0.796 cm^2 for CSA, and 0.603 – 0.937 AU for EI. Yet these were all relatively small expressed relative to the mean as indicated in Table 3. The absolute SEMs are similar to previous works, in which VL CSA had a SEM of 2.12 cm^2 [7] or a SEM of 0.410 cm^2 [25]. Similarly, RF CSA has been shown to have a SEM of 0.282 cm^2 [25]. The low MD values found in the RF and VL outcomes in this study were similar to the VL MD reported by Ahtiainen et al. of 1.1 cm^2 for CSA [5]. However, it is worth noting that there are multiple ways to calculate the SEM. The present study equated SEM as the MS_E from the ANOVA results, as suggested by Weir [13] and Hopkins [24]. Sahinis et al. calculated the SEM as the SD multiplied by the square root of 1 – the ICC [25], while Rosenberg et al. used the same equation as the present study [26]. Using the present study's data as an example, our SEM value (calculated from the square root of the MS_E) for the MT of the VL was 0.796 cm, but if calculated as the square root of 1 – $ICC_{2,k}$ the SEM would then be 1.144 cm. Although these differences may not occur in each situation or might not severely impact the interpretation of the SEM (e.g., if the SEM is presented as a percentage of the mean), it is still pertinent to acknowledge that there are deviations in how reliability statistics can be calculated.

To the authors' knowledge, this is the first study to assess reliability of the panoramic ultrasound measurements of the FDI. Overall, the inter-rater reliability of the FDI ranged from $ICC_{2,k}$ values of 0.614 for MT, 0.842 for CSA, and 0.962 for EI. This could be due to the fact that the FDI is in an uneven body region on the hand, which may limit the reliability of FDI image analysis [10]. A previous study utilized B-mode ultrasound to examine the reliability of CSA measurements of the FDI on two separate occasions. An ICC of 0.979 was identified by Martin et al. [27]. However, along with the differences in ultrasound technique (panoramic v. B-mode), Martin et al. utilized a different ultrasound device [27] than the present study [20] (Phillips v. GE, respectively). Lastly, the CSA of the FDI in the Martin et al. study was estimated from the diameter of the muscle determined in the ultrasound device's software [27]. With these methodological differences and the resulting ICCs determined from the analyses of these two studies, it is difficult to compare results. Therefore, it is imperative to continue attempt to standardize ultrasound image acquisition and analysis for skeletal muscles in order to easily compare results across laboratory groups, populations, and studies.

For intra-rater reliability, $ICC_{2,k}$ ranged from 0.961 to 0.999. However, significant systematic variability was detected in the EI of the VL for Rater 1, the CSA and MT of the VL for Rater 2, and the CSA and EI of the VL and RF, respectively, for Rater 3. Similar to the significant systematic variability in inter-rater reliability, these significant findings may be due to small mean differences [13]. However, the SEM and MD values within each rater were small, similar to the inter-rater SEM and MD. Tomko et al. measured CSA of the RF within the same rater on two separate days and demonstrated $ICC_{2,k}$ values of 0.98 [28]. For the VL CSA, an $ICC_{2,1}$ of 0.87 was observed by Melvin et al. [7]. Each rater in the present study had varying levels of experience with using ImageJ. However, reliable analyses were observed and this supports the findings of Zaidman et al., in which it was determined that minimal training was necessary to obtain highly reliable gray-scale intensities of the elbow flexors and RF [29]. As our ICCs within each rater across three separate analyses, it would appear that our instructions on ultrasound image analysis allowed each rater to reliably analyze the selected outcomes.

Conclusion

In conclusion, the results of the present study suggest that panoramic ultrasound is a reliable methodology to assess MT, CSA, and EI of the VL, RF, and FDI across multiple raters. The use of three separate raters was unique and novel in this present study. Although systematic variability was present in some of the selected muscles and outcomes, the reliability was still high. The FDI had the lowest reliability of the three muscles and this was most likely due to its unique location on the hand. However, this study was the first to assess ultrasound reliability of the FDI and hopefully can provide further insight into ultrasound analysis of the skeletal muscles of the hand. Future studies should continue to assess the impact of raters with varying levels of experience, ultrasound training modalities, ultrasound devices, analysis software, and analysis instructions on the reliability of ultrasound quantification of skeletal muscle characteristics.

Abbreviations

ANOVA = analysis of variance

CSA = cross-sectional area

CT = computed tomography

EI = echo intensity

FDI = first dorsal interosseus

ICC = intra-class correlation coefficient

MD = minimal difference needed to be considered real

MRI = magnetic resonance imaging

MT = muscle thickness

NIH = National Institutes of Health

RF = rectus femoris

SEM = standard error of the measurement

VL = vastus lateralis

Declarations

Author contributions. AH conceptualized and designed the study. CC, ON, KY, and AH collected and analyzed the data. CC wrote up the manuscript which was revised by AH. All authors read and approved the final manuscript.

Conflicts of interest. No conflicts of interest or competing interests, financial or otherwise, are declared by the authors.

Funding. Not applicable.

Data availability. Data and materials can be requested through the corresponding author.

Code availability. Not applicable.

References

1. Palmer TB, Akehi K, Thiele RM, et al (2015) 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 41:675–684. https://doi.org/10.1016/j.ultrasmedbio.2014.10.011
2. Scott JM, Martin DS, Ploutz-Snyder R, et al (2012) Reliability and validity of panoramic ultrasound for muscle quantification. Ultrasound Med Biol 38:1656–1661. https://doi.org/10.1016/j.ultrasmedbio.2012.04.018
3. Yang J-H, Eun S-P, Park D-H, et al (2019) The Effects of Anterior Cruciate Ligament Reconstruction on Individual Quadriceps Muscle Thickness and Circulating Biomarkers. Int J Environ Res Public Health 16:4895. https://doi.org/10.3390/ijerph16244895
4. Filippo M, Lars A-N, Maria S, Sandra A-B (2019) Inter-rater and intra-rater reliability of ultrasound imaging for measuring quadriceps muscle and non-contractile tissue thickness of the anterior thigh. Biomed Phys Eng Express 5:037002. https://doi.org/10.1088/2057-1976/ab102f
5. Ahtiainen JP, Hoffren M, Hulmi JJ, et al (2010) Panoramic ultrasonography is a valid method to measure changes in skeletal muscle cross-sectional area. Eur J Appl Physiol 108:273–279. https://doi.org/10.1007/s00421-009-1211-6
6. Franchi MV, Longo S, Mallinson J, et al (2018) Muscle thickness correlates to muscle cross-sectional area in the assessment of strength training-induced hypertrophy. Scand J Med Sci Sports 28:846–853. https://doi.org/10.1111/sms.12961
7. Melvin MN, Smith-Ryan AE, Wingfield HL, et al (2014) Evaluation of muscle quality reliability and racial differences in body composition of overweight individuals. Ultrasound Med Biol 40:1973–1979. https://doi.org/10.1016/j.ultrasmedbio.2014.03.012
8. Scott JM, Martin DS, Ploutz-Snyder R, et al (2017) Panoramic ultrasound: a novel and valid tool for monitoring change in muscle mass. J Cachexia Sarcopenia Muscle 8:475–481.

https://doi.org/10.1002/jcsm.12172

9. Valera-Calero JA, Ojedo-Martín C, Fernández-de-las-Peñas C, et al (2020) Reliability and Validity of Panoramic Ultrasound Imaging for Evaluating Muscular Quality and Morphology: A Systematic Review. Ultrasound Med Biol. https://doi.org/10.1016/j.ultrasmedbio.2020.10.009
10. Noorkoiv M, Nosaka K, Blazevich AJ (2010) Assessment of quadriceps muscle cross-sectional area by ultrasound extended-field-of-view imaging. Eur J Appl Physiol 109:631–639. https://doi.org/10.1007/s00421-010-1402-1
11. Bemben MG (2002) Use of diagnostic ultrasound for assessing muscle size. J Strength Cond Res 16:103–108
12. Valera-Calero JA, Sánchez-Jorge S, Álvarez-González J, et al (2020) Intra-rater and inter-rater reliability of rehabilitative ultrasound imaging of cervical multifidus muscle in healthy people: Imaging capturing and imaging calculation. Musculoskelet Sci Pract 48:102158. https://doi.org/10.1016/j.msksp.2020.102158
13. Weir JP (2005) Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res 19:231–240
14. Teyhen DS, George SZ, Dugan JL, et al (2011) Inter-Rater Reliability of Ultrasound Imaging of the Trunk Musculature Among Novice Raters. J Ultrasound Med 30:347–356. https://doi.org/10.7863/jum.2011.30.3.347
15. Fellows SJ, Noth J, Schwarz M (1998) Precision grip and Parkinson’s disease. Brain 121:1771–1784. https://doi.org/10.1093/brain/121.9.1771
16. Sailer A, Molnar GF, Paradiso G, et al (2003) Short and long latency afferent inhibition in Parkinson’s disease. Brain 126:1883–1894. https://doi.org/10.1093/brain/awg183
17. Pillen S (2010) Skeletal muscle ultrasound. Eur J Transl Myol 20:145–156. https://doi.org/10.4081/ejtm.2010.1812
18. Lanferdini FJ, Manganelli BF, Lopez P, et al (2019) Echo Intensity Reliability for the Analysis of Different Muscle Areas in Athletes. J Strength Cond Res 33:3353–3360. https://doi.org/10.1519/JSC.0000000000003063
19. Pillen S, van Alfen N (2011) Skeletal muscle ultrasound. Neurol Res 33:1016–1024. https://doi.org/10.1179/1743132811Y.0000000010
20. Miller JD, Sterczala AJ, Trevino MA, Herda TJ (2018) Examination of muscle composition and motor unit behavior of the first dorsal interosseous of normal and overweight children. J Neurophysiol 119:1902–1911. https://doi.org/10.1152/jn.00675.2017
21. Burton AM, Stock MS (2018) Consistency of novel ultrasound equations for estimating percent intramuscular fat. Clin Physiol Funct Imaging 38:1062–1066. https://doi.org/10.1111/cpf.12532
22. Shrout PE, Fleiss JL (1979) Intraclass correlations: Uses in assessing rater reliability. Psychol Bull 86:420–428. https://doi.org/10.1037/0033-2909.86.2.420

23. Reis FJJ dos, Silva V de B e, Lucena RN de, et al (2016) Measuring the Pain Area: An Intra- and Inter-Rater Reliability Study Using Image Analysis Software. Pain Pract 16:24–30. https://doi.org/10.1111/papr.12262
24. Hopkins WG (2000) Measures of reliability in sports medicine and science. Sports Med Auckl NZ 30:1–15. https://doi.org/10.2165/00007256-200030010-00001
25. Sahinis C, Kellis E, Galanis N, et al (2020) Intra- and inter-muscular differences in the cross-sectional area of the quadriceps muscles assessed by extended field-of-view ultrasonography. Med Ultrason 22:152–158. https://doi.org/10.11152/mu-2302
26. Rosenberg JG, Ryan ED, Sobolewski EJ, et al (2014) Reliability of panoramic ultrasound imaging to simultaneously examine muscle size and quality of the medial gastrocnemius. Muscle Nerve 49:736–740. https://doi.org/10.1002/mus.24061
27. Martin D, Cooper S, Sale C, et al (2015) Reliability of force per unit cross-sectional area measurements of the first dorsal interosseus muscle. J Sports Sci 33:1159–1165. https://doi.org/10.1080/02640414.2014.986504
28. Tomko PM, Muddle TW, Magrini MA, et al (2018) Reliability and differences in quadriceps femoris muscle morphology using ultrasonography: The effects of body position and rest time. Ultrasound Leeds Engl 26:214–221. https://doi.org/10.1177/1742271X18780127
29. Zaidman CM, Wu JS, Wilder S, et al (2014) Minimal training is required to reliably perform quantitative ultrasound of muscle. Muscle Nerve 50:124–128. https://doi.org/10.1002/mus.24117

Figures

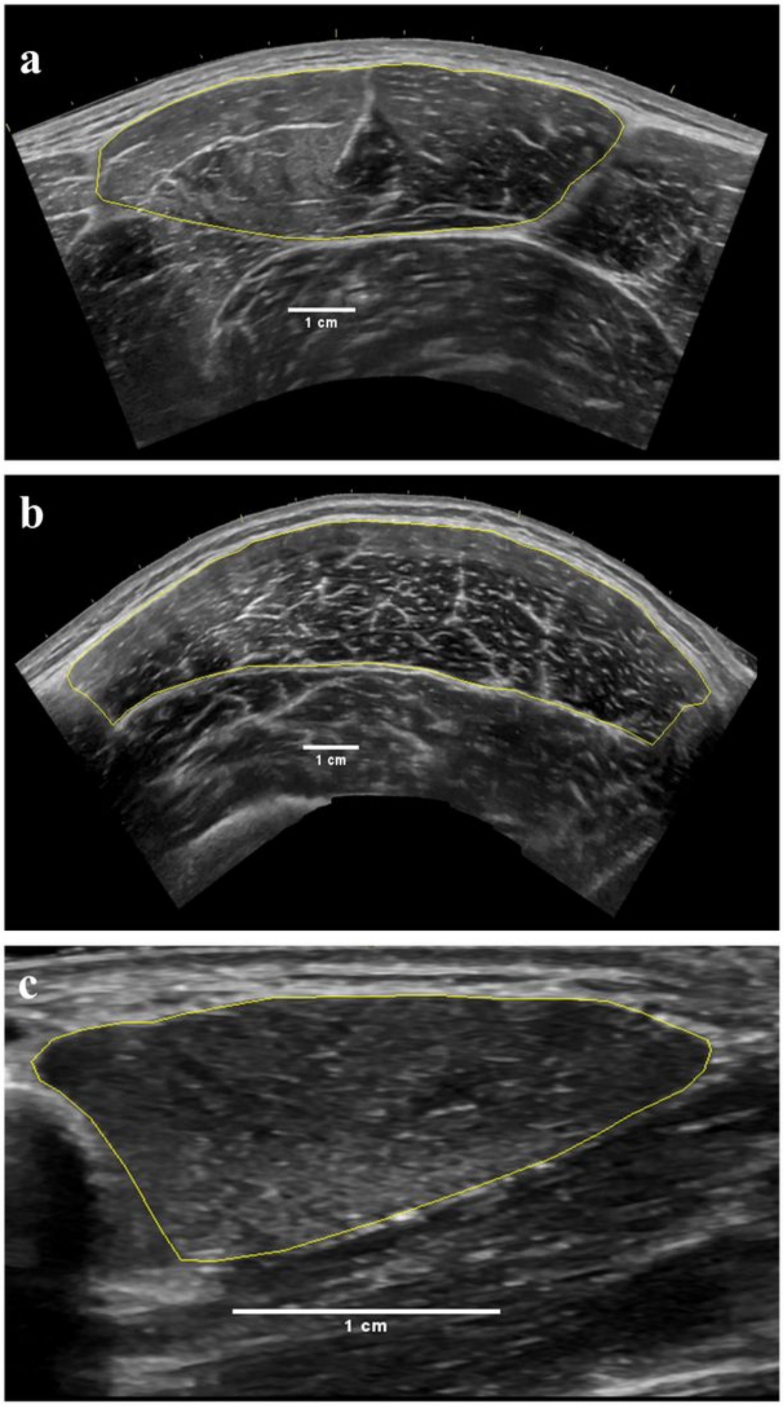

Figure 1

ImageJ of the three skeletal muscles utilized in the analyses; 1a = rectus femoris; 1b = vastus lateralis; 1c = first dorsal interosseus.