Research Square

Preprints are preliminary reports that have not undergone peer review. They should not be considered conclusive, used to inform clinical practice, or referenced by the media as validated information.

Intra- and Inter-Rater Reliability of Rectus Femoris Muscle Quality Measures Using Ultrasound Imaging: Novice vs Experienced Rater

Isabel N. King

ikingO511@outlook.com

University of the Sunshine Coast

Jacqueline Roots

Queensland University of Technology

James Newman

University of the Sunshine Coast

Grace L. Rose

University of the Sunshine Coast

Christopher D. Askew

University of the Sunshine Coast

Mia A. Schaumberg

University of the Sunshine Coast

Robert Buhmann

University of the Sunshine Coast

Research Article

Keywords: echogenicity, ultrasonography, muscle quality, older adults, reliability

Posted Date: October 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11176776/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

The assessment of leg muscle quality using ultrasound imaging has emerged as a valuable tool for investigating age-related changes, functional limitations, and health outcomes. However, practical challenges associated with the availability and cost of accredited sonographers in research settings underscore the need to explore the proficiency of less experienced operators utilising mobile ultrasound devices while evaluating intra- and inter-rater reliability. This study aimed to evaluate and compare the intra- and inter-rater reliability of rectus femoris muscle quality measurements (muscle thickness, adipose thickness and echo intensity) in healthy older adults, conducted by novice and experienced raters.

Methods

Twenty healthy older adults (eight women, 12 men, mean age 71.6 ± 4.7 years) were included. Ultrasound images of the left rectus femoris were obtained with participants in the supine position. Measurements were performed by two novice raters and one experienced sonographer. Intra-class correlation coefficients (ICCs) were calculated to evaluate intra-rater and inter-rater reliability of echo intensity and muscle thickness measures. T-tests were conducted to assess differences between individual raters (experienced rater vs. Novice1; experienced rater vs. Novice2; Novice1 vs. Novice2), and standard error of the mean (absolute and as a percentage of the mean).

Results

The intra-rater ICC values for all measurements for the experienced rater (ICC all > 0.97) and for novices (ICC all > 0.96) were excellent. The inter-rater reliability ICC values showed good to excellent reliability (0.76–0.967) between novice and experienced raters for all measures. Bland-Altman plots showed no systematic errors in novice vs. expert-obtained measures, confirming novice assessment reliability. Variation within individual raters across trials was minimal, with coefficients of variation ranging from 1.53% to 4.22% and standard errors of the mean ranging from 0.15 to 0.62% for all variables.

Conclusions

Our findings support using ultrasound imaging to assess morphological leg muscle quality in older adults, indicating its potential as a reliable and practical tool for routine assessments in various research and healthcare settings. The observed agreement between novice and experienced raters suggests that with appropriate training, novice raters can effectively contribute to muscle quality assessments in research contexts.

Background

Examining muscle quality (MQ) among older individuals is important for evaluating the ageing process and its impact on physical functionality (1–4). Muscle quality, delineated as muscle tissue's structural and compositional attributes, is characterised by muscle size, architecture, and intramuscular adiposity (5). These characteristics directly influence the functional capacity of muscles (1, 6, 7). Age-related declines, particularly in lower limb muscle mass and quality (8, 9), correlate with impaired mobility, an elevated propensity for falls, and a decline in quality of life (10–12). In this regard, assessing lower limb MQ in older adults is crucial as it may provide valuable prognostic insights into age-related functional decline and limitations (13–17). Such information can help practitioners design targeted interventions to improve mobility, mitigate fall risks, and foster health and autonomy in the ageing demographic.

Ultrasound (US) devices have emerged as a commonly used modality for assessing MQ in older adults (18), and the widespread availability of portable devices has enhanced accessibility for use in research. Muscle US assessment typically includes several morphological properties, including muscle thickness (MT), adipose thickness (AT) and echo intensity (EI) (19). Muscle thickness and EI are pivotal in evaluating MQ, especially in older adults. While MT provides valuable insights into muscle size and atrophy (20), EI provides insight into muscle composition, with higher values indicative of increased intramuscular fat and fibrosis, both of which are linked to diminished muscle function and mobility among elderly individuals (20, 21).

While prior research has demonstrated good intra-rater reliability for US measurements of the rectus femoris muscle (22) and the anterior thigh (23), no studies have compared the reliability of rectus femoris MT and EI measurements between novice and experienced raters in healthy older adults. Several factors influence the reliability of US measures, including the operator's level of experience (24), between-day fluctuations in EI (25), and the method employed (26) (e.g., variations in scanning techniques and protocols may impact the accuracy and consistency of the results). Additionally, degradation of MQ can make US assessments difficult for novice investigators. Considering the growing availability of portable US devices, which may encourage their use by researchers with insufficient training, the reliability of these measures compared with expert operators must be assessed.

Examining the inter- and intra-rater reliability of muscle US measures between novice and experienced raters is imperative for upholding the precision and consistency of MQ assessments in research and clinical settings. A better understanding of the ability of novice raters to evaluate MQ using portable US devices reliably would potentially enhance the accessibility and support the broader use of these methods in future research. Therefore, this study aims to investigate and compare the inter- and intra-rater reliability of US-derived measures of rectus femoris muscle thickness, adipose tissue, and echo intensity, drawing comparisons between experienced and novice raters.

Methods

Study overview

A subset of 20 participants was randomly selected from a larger cohort of an exercise study to assess intra- and inter-reliability in the measurements of EI, MT and AT of the left rectus femoris at one assessment timepoint during their exercise trial (Week 0, Week 6, or Week 13). All participants underwent an anthropometric (height, weight, waist circumference) and body composition assessment. Body mass index (BMI) was calculated by dividing body mass by the square of height (kg/m^2). Dual-energy x-ray absorptiometry (Lunar iDXA, GE Healthcare, Madison, WI, USA) was used for body composition measurements, including fat-free mass, percentage body fat and leg lean mass. All US assessments were conducted immediately after the DXA scans as the participants were supine and rested for about 10 minutes. Two novice raters (Novice1 and Novice2) performed one assessment each, comprising three scans and six pictures by the same novice operator. One experienced rater (Sonographer with > 10 years of experience; Expert) performed two US assessments, separated by 10 minutes, comprising six images per assessment. Assessments followed the same order: Expert - Novice1- Novice2 – Expert. Prior to the study, both novice raters had two-hour training sessions with the Expert until the experienced user was satisfied with their technique. All raters were blinded to each other's measurements - individual identifiers were assigned to US scans, and a physical curtain was used to ensure separation between scanners during the scanning process. Independent image analysis was conducted to prevent bias in the interpretation of results. The study was approved by the University of the Sunshine Coast Human Research Ethics Committee (approval number S221785) and conducted following the principles of the Declaration of Helsinki. Participants were only included after providing signed written informed consent.

Participants

Healthy older adults (n = 20 [n = 12 men, n = 8 women], average age 71.6 ± 4.7 years) with no known pre-existing medical conditions (e.g., cardiac conditions, mental illness, cognitive impairment, poor glucose regulation) were included in this study. Participants were excluded from the study if they reported any acute or chronic musculoskeletal injury of the lower limb.

Sample size (n = 20) was determined using an estimation-based approach (27) and calculated assuming a true intra-class correlation coefficient (ICC) of 0.9 (22), a 95% confidence interval (CI) width around the ICC of 0.2 and a desired precision of 80%. The number of raters/repeats was set to three when calculating the sample size, as three raters were included in our study.

Ultrasound

Ultrasound images were obtained on a Butterfly iQ device (Butterfly Network Inc., Burlington, MA, USA) on the MSK- Soft Tissue pre-set. The imaging parameters were consistent in all acquisitions (frequency: 5 MHz, gain: 60, focal distance kept at factory settings), depth was optimised for each participant. The

RF muscle of the left leg was identified on B-mode and scanned along its length in an axial orientation. Rectus femoris MT was measured with the US probe placed on the anterior surface of the thigh, at the midpoint of the length between the anterior superior iliac spine and the upper border of the patella, referring to the measurement position of Pardo et al (28). The probe was maintained in this position on the skin surface, and minimal contact pressure was applied during the measurements to obtain good-quality images. Images were optimised for measuring MT/AT and EI. Intra-rater reliability for novice raters (Novice1 and Novice2) was assessed by comparing measurements across three successive scans performed by each novice on the same subject. The two measurements conducted by the Expert were used for intra-rater reliability. Novice1 and Novice2 were compared to Expert rater measurements to assess inter-rater reliability (29).

Image Assessment

Muscle thickness, both with and without the epimysium (MT and MTex, respectively), and the thickness of the adipose layer (AT), were measured using the same image. Mean MT/MTex (cm) was determined from three images (Fig. 1.). The EI of the muscle was measured by grayscale analysis using ImageJ software (Version 1.54D, National Institutes of Health, USA). The region of interest (ROI) was manually traced and included all tissue within the epimysium. The EI values were determined as the corresponding index of MQ ranging between 0 and 255 arbitrary units (A.U.) (black = zero, white = 255), where higher EI values indicate a greater amount of intramuscular fat and connective tissue (30) (Fig. 2).

Insert Fig. 1 **Here**

Insert Fig. 2 **Here**

Statistical analysis

Data were analysed using Microsoft 365 Excel and SPSS (version 22.0, SPSS, Inc., Chicago, IL, USA) statistical software package. Means and standard deviations for participant characteristics were calculated for each variable (Table 1). Histograms and the Shapiro-Wilks test were used to determine whether data were normally distributed. Means and standard deviations are reported for each rater and variable, and repeated measures ANOVA was used to test for significant differences within raters **(**Table 2**).** T-tests were used to determine if there were differences between individual raters (ER vs. Novice1; ER vs. Novice2; Novice1 vs. Novice2). Reliability was assessed using several measures, including the intraclass correlation coefficient (ICC), coefficient of variation and standard error of the mean (both absolute and as a percentage of the mean). The ICC was calculated using a two-way random effects model:

$$\frac{\mathrm{MSS} - \mathrm{MSE}}{\mathrm{MSS}\,\frac{\mathrm{k(MST} - \mathrm{MSS)}}{\mathrm{n}}}$$

where MS_S = participant mean square, MS_E = error mean square, MS_T = total mean square, k = number of repeat measurements/raters and n = number of participants.

The coefficient of variation (%CV) was calculated as:

$$\%CV = (SD \div mean) * 100$$

The standard error of the measurement was calculated as:

$$SEM = SD * \sqrt{(1 - ICC)}$$

Where SD = the standard deviation of the measurements.

ICCs were interpreted as follows: < 0.5 = poor reliability; 0.50 to 0.75 = moderate reliability; 0.75 to 0.90 = good reliability; > 0.90 excellent reliability (31). We also present Bland-Altman plots to demonstrate that differences between the expert and novices are randomly distributed across mean values.

Results

Participant characteristics (n = 20, 12 male, eight female) are presented in Table 1.. Participants had an average body mass of 79.8 ± 17.7 kg and fell within a normal to overweight BMI range (23.6–32.6 $kg.m^{-2}$). Average sit-to-stand count was 19.5 ± 5.3.

Table 1
Participant baseline characteristics (mean ± SD)

n	**20**
% female	40 (n = 8)
Age (years)	71.6 ± 4.7
Body mass (kg)	79.8 ± 17.7
Height (cm)	167.9 ± 9.9
BMI ($kg.m^{-2}$)	28.1 ± 4.5
FFM (kg)	55.7 ± 12.1
FM (kg)	24.1 ± 9.0
BF%	29.8 ± 7.0
LLM (Kg)	16.6 ± 4.2

Descriptive characteristics of participant characteristics derived from DXA, presented as mean ± SD. BF%: Body fat percentage, BMI: Body mass index, FFM: fat-free mass, FM: fat mass, LLM: Leg lean mass

Intra-Rater and Inter-Rater Reliability

Intra-rater reliability measures are presented in Table 2.. Intra-rater reliability was excellent for all measures (ICCs = 0.96 to 0.99). There were no differences between trials within individual raters for any variable. Variation between trials within individual raters was small (%CV = 1.53 to 4.22%; %SEM = 0.15 to 0.62%) for all variables.

Inter-rater reliability measures are presented in Table 3. There were no significant differences between measures between the experienced and novice raters for any variables. Inter-rater reliability was excellent for MT (ICC = 0.92, 95% CI = 0.80 to 0.97) and AT (ICC = 0.95, 95% CI = 0.92 to 0.96) and good for MTex (ICC = 0.87, 95% CI = 0.67 to 0.95). Inter-rater reliability for EI was moderate (ICC = 0.66, 95% CI = 0.34 to 0.87). Notably, the ER vs. Novices %SEM values were consistently higher for inter-rater comparisons than intra-rater assessments across all measures, emphasising the greater variability introduced by different raters for MT.

Table 2
Intra-rater reliability of US measurements of the left RF muscle in older adults between Expert rater (Expert) and each novice (Novice1, Novice2).

Reliability	Mean ± SD	%CV	ICC (95% CI)	SEM (%)
Measure	**all (n = 20)**			
MT (cm)				
ER	2.41 ± 0.40	1.53	0.99 (0.98–0.99)	0.004 (0.17)
Novice1	2.27 ± 0.39	2.30	0.99 (0.98–0.99)	0.005 (0.22)
Novice2	2.31 ± 0.34	2.94	0.97 (0.94–0.99)	0.012 (0.52)
MTex (cm)				
ER	2.28 ± 0.38	1.93	0.99 (0.97–0.99)	0.005 (0.22)
Novice1	2.06 ± 0.38	2.63	0.99 (0.97–0.99)	0.006 (0.29)
Novice2	2.11 ± 0.32	3.22	0.97 (0.94–0.99)	0.011 (0.52)
AT (cm)				
ER	1.46 ± 0.59	3.55	0.99 (0.99–0.99)	0.003 (0.21)
Novice1	1.32 ± 0.55	2.36	0.99 (0.99–0.99)	0.002 (0.15)
Novice2	1.32 ± 0.60	4.22	0.98 (0.97–0.99)	0.007 (0.53)
EI (A.U.)				
ER	69.17 ± 11.61	3.50	0.97 (0.95–0.99)	0.370 (0.53)
Novice1	74.40 ± 13.70	3.90	0.97 (0.94–0.99)	0.460 (0.62)
Novice2	72.46 ± 9.26	2.66	0.96 (0.92–0.98)	0.358 (0.49)

%CV: Coefficient of variation; ICC: intraclass correlation coefficient, ICC 95% CI: 95% confidence interval for ICC; SEM standard error of measurement; SEM (%): standard error of measurement expressed as percentage of the mean. MT: muscle thickness with epimysium; MTex: muscle thickness without epimysium; AT: adipose thickness; EI: echo intensity.

Table 3
Inter-rater reliability of the rectus femoris ultrasound measurements between expert and novice raters.

MT (cm)	Expert	Novice 1	Novice 2	ER vs Novice1[1]	ER vs Novice2[1]	%CV[2]	ICC (95% CI)[3]	SEM (%)[4]
	2.41 ± 0.40	**2.27 ± 0.39**	**2.31 ± 0.34**	**0.25**	**0.40**	**3.54**	**0.92 (0.80–0.97)**	**0.02 (0.8)**
MTex (cm)	2.28 ± 0.38	2.06 ± 0.38	2.11 ± 0.32	0.08	0.14	5.10	0.87 (0.67–0.95)	0.04 (1.8)
AT (cm)	1.46 ± 0.59	1.32 ± 0.55	1.32 ± 0.60	0.46	0.418	7.82	0.95 (0.92–0.96)	0.02 (1.7)
EI (A.U)	69.17 ± 11.61	74.40 ± 13.70	72.46 ± 9.26	0.21	0.34	6.64	0.66 (0.34–0.87)	2.73 (3.8)

,[1]P-value: type 1 error rate simple T-test comparing ER and Novices [2]%CV: Coefficient of variation between ER and Novice1, Novice2 averaged, [3]ICC: intraclass correlation coefficient calculated with a one-way random effects model and reported as the average between ER vs Novice1 and ER vs Novice2, 95% CI: 95% confidence interval,; [4]SEM: standard error of measurement and reported as the average between ER vs Novice1 and ER vs Novice2; SEM (%): standard error of measurement expressed as percentage of the mean of all three raters.

Systematic Error in thickness and echo intensity

Bland-Altman plots demonstrating differences in echo intensity and muscle thickness measures between novices and the expert are displayed in Fig. 3. These figures demonstrate a random distribution of differences across mean values for both measures, suggesting there are no systematic differences between novices and the expert for either measure.

Insert Fig. 3 **Here**

Discussion

This study assessed the intra- and inter-rater reliability of ultrasound measures of muscle quality (MT, MTex, AT, and EI) using a portable device in 20 healthy older adults. Intra-rater ICCs for all raters and measurements were excellent (all > 0.96). When comparing novice and experienced raters, inter-rater reliability ranged from moderate (echo intensity) to excellent (muscle thickness and adipose thickness) Additionally, Bland-Altman plots revealed only a small mean difference between novice and experts for

key measures, further highlighting the consistency of intra-rater assessments. ICC values were good to excellent for all measures except EI, where reliability was moderate. Notably, the intra-rater %CV was smaller (0.82–4.16%) than the %CVs for inter-rater (4.84–9.49%), indicating greater variability between raters than within raters. Importantly, establishing reliability allows for valid and precise measures, facilitating research to better understand the impact of aging, explore associations with risk factors, and determine the magnitude of clinical effects.

In contrast to Sobolewski et al. (32), we observed a higher inter-rater ICC for EI (ICC = 0.66, 95% CI = 0.34 to 0.87) and an intra-rater ICC ranging from 0.96 to 0.97. The disparity in findings may arise from differences in US methodology and participant characteristics. Sobolewski and colleagues (32) employed panoramic ultrasound, capturing broader muscle areas; our images were acquired at a singular point along the muscle belly in the transverse plane. Panoramic ultrasound captures a wider muscle area, which undergoes automated integration by the machine. This process, while mechanised, may introduce complexities and potential errors, contributing to heightened variability between images and potentially explaining the lower intra-rater reliability observed by Sobolewski et al. (32). Additionally, our study focussed on older adults, contrasting with Sobolewski and colleagues' participants, who had a mean age of 20.5 years. Age-related changes in MQ could also contribute to increased variability in image interpretation, particularly for novice raters unfamiliar with imaging muscle using ultrasound. Factors such as potential learning effects and fatigue, as noted by Castro et al. (33) and Sions et al. (34), may also contribute to variability in repeated measurements by the same rater.

For MT ultrasound measures, Takahashi et al. (22) reported excellent intra-rater reliability (ICC = 0.95, 95% CI 0.86 to 0.99) and moderate inter-rater reliability (ICC = 0.70, 95% CI 0.23 to 0.90). Similarly, Pinto-Ramos and colleagues (35) report excellent ICC values for both inter-rater (0.91, 95% CI = 0.86 to 0.96) and intra-rater (0.96, 95% CI = 0.94 to 0.97) reliability. While our results are similar to previously published work, a discrepancy exists between the inter-rater ICC we observed for MT and the ICC reported by Takahashi et al. (36). This discrepancy suggests potential variations in the methodologies employed, participant characteristics, or both. This difference hints at potential variations in the methodologies or participant characteristics. Notably, Takahashi et al. (36) employed similar imaging methods, but the absence of a portable device and the calculation of MT 'live' during image acquisition, along with one week between measurements, could contribute to increased variability in their study. The recruitment of exclusively young males might mitigate variability, suggesting that the observed difference in reliability could be attributed to diverse devices and the extended time between measurements.

Interpreting EI often involves subjective judgement influenced by visual assessment. Differences in perceptual skills and experience between expert and novice raters could contribute to evaluation variability. The inherent complexity of interpreting EI, involving subtle variations in muscle tissue echogenicity, may introduce ambiguity and increase variability. Methodological aspects, such as inconsistent scanning techniques or variations in region of interest selection, could also contribute to increased variability, emphasising the importance of standardisation across studies. This holds significant implications for interpreting results, as EI is a key metric for assessing MQ, and poor reliability

could lead to incorrect conclusions. Future research should aim to understand the factors underpinning poor inter-rater reliability of EI.

A notable strength of our study is the use of two novice raters to evaluate inter-rater reliability. By employing novice raters, we ensured a thorough assessment of the consistency and accuracy of measurements, allowing for a comprehensive examination of potential sources of variability.

The current study has some limitations that require consideration. Firstly, this study did not assess the degree of inward pressure and the angle at which the transducer was applied to the leg during measurements (37). This raises the possibility that inter-rater reliability was influenced by technical factors, potentially explaining the lower levels of reliability in EI compared with other variables. Angling and transducer pressure were minimised whilst achieving the best quality image. However, there is currently no widely accepted method available for minimising the effects of inward transducer pressure; it could be helpful to establish a standardised protocol with precise transducer conditions during data collection, consider US guidance, particularly in novice raters, and conduct post-study analyses to minimise their impact on measurement outcomes. Furthermore, the likelihood of patient movement or leg muscle contractions (motion artefacts) during the assessment introduces another source of measurement variability (38). This should also be considered when interpreting the results, as it could impact the reliability of all MQ measurements employed in this study. To minimise participants' movement or leg muscle contractions during the assessments, participants were positioned comfortably and real-time feedback from the ultrasound device was used to minimise muscle contraction. While the assessments in the current study were not randomised, this does not necessarily present a limitation. However, it is essential to acknowledge potential sources of bias from non-randomised assessments and consider ways to address them in future research. Additionally, the focus on older adults in the study may introduce age-related factors that influence MQ and could limit the generalisability of the results to other age groups (34). Moreover, it's crucial to consider whether participant characteristics, such as age-associated muscle atrophy and fatty infiltration, may have influenced reliability in any way. This consideration could provide valuable insights into the factors influencing variability in ultrasound measurements and inform strategies for improving reliability, particularly when novice assessors are involved. Understanding how participant characteristics interact with the assessment process can enhance the interpretation of study findings and guide future research efforts to refine measurement protocols and minimise sources of variability.

Conclusions

In conclusion, in our study, novice raters used a portable US device to produce comparable US measurements of muscle quality to an expert rater, indicating the potential for consistent image reproduction and derived variables within the RF of older adults. The observed excellent intra-rater reliability across all raters and the absence of systematic errors, as indicated by the random distribution of differences in echo intensity and muscle thickness measures between novices and the expert in Bland-Altman plots, underscores the reliability of novice-obtained measurements relative to those

obtained by the expert. Nonetheless, there was variation between different raters when assessing EI, suggesting potential challenges in achieving consistent results in this specific parameter. Our findings demonstrate that portable ultrasound devices yield reliable results when used by novice raters, increasing the feasibility of US in a wider range of research settings. Further work should aim to understand the factors underpinning lower levels of reliability in echo intensity measures.

Abbreviations

AT
adipose thickness
A.U.
arbitrary units
BF%
body fat percentage
BMI
body mass index
CI
confidence interval
%CV
coefficient of variation
DXA
dual-energy X-ray absorptiometry
EI
echo intensity
FFM
fat-free mass
FM
fat mass
ICC
intraclass correlation coefficient
LLM
leg lean mass
MQ
muscle quality
MT
muscle thickness (with epimysium)
MTex
muscle thickness (without epimysium)
RF
rectus femoris

ROI
region of interest
SEM
standard error of measurement
US
ultrasound

Declarations

Ethics approval and consent to participate

This study was approved by the University of the Sunshine Coast Human Research Ethics Committee (approval number S221785) and was conducted in accordance with the principles of the Declaration of Helsinki. All participants provided signed written informed consent prior to participation.

Consent for publication

Not applicable. The ultrasound images presented do not contain any information that could identify individual participants.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was supported by a UniSC LAUNCH Grant (grant number: 980027690).

Author Contribution

INK contributed to conceptualisation, data collection, data analysis and interpretation, and drafting of the manuscript. RB and MAS contributed to conceptualisation, methodology and supervision. JR (experienced rater/sonographer) contributed to data acquisition and methodology. JN and GLR contributed to data collection and analysis. CDA contributed to conceptualisation and critical revision of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Not applicable

Data Availability

The datasets generated and analysed during the current study are not publicly available because the conditions of ethics approval granted by the University of the Sunshine Coast Human Research Ethics Committee (approval number S221785) and the consent provided by participants do not permit open sharing of individual participant data. De-identified data are available from the corresponding author (Isabel N. King, iking0511@outlook.com) on reasonable request.

References

1. Straight CR, Brady AO, Evans EM. Muscle Quality in Older Adults: What Are the Health Implications? Am J Lifestyle Med. 2015;9(2):130–6. 10.1177/1559827613510681.
2. Sui SX, Williams LJ, Holloway-Kew KL, Hyde NK, Leach S, Pasco JA. Associations Between Muscle Quality and Cognitive Function in Older Men: Cross-Sectional Data From the Geelong Osteoporosis Study. J Clin Densitom. 2021;S1094695021000251. 10.1016/j.jocd.2021.03.007.
3. Van Den Broeck J, Héréus S, Cattrysse E, Raeymaekers H, De Maeseneer M, Scafoglieri A. Reliability of Muscle Quantity and Quality Measured With Extended-Field-of-View Ultrasound at Nine Body Sites. Ultrasound Med Biol. 2023;49(7):1544–9. 10.1016/j.ultrasmedbio.2023.02.018.
4. Wen QR, Wu M, Liu Q, Lyu J, Guo Y, Bian Z, et al. [Correlation between chronic diseases and low muscle mass, strength and quality in adults in China]. Zhonghua Liu Xing Bing Xue Za Zhi Zhonghua Liuxingbingxue Zazhi. 2021;42(11):1948–54. 10.3760/cma.j.cn112338-20200910-01146 . PubMed PMID: 34818839.
5. Naimo MA, Varanoske AN, Hughes JM, Pasiakos SM. Skeletal Muscle Quality: A Biomarker for Assessing Physical Performance Capabilities in Young Populations. Front Physiol. 2021;12:706699. 10.3389/fphys.2021.706699.
6. Robles PG, Sussman MS, Naraghi A, Brooks D, Goldstein RS, White LM, et al. Intramuscular Fat Infiltration Contributes to Impaired Muscle Function in COPD. Med Sci Sports Exerc. 2015;47(7):1334–41. 10.1249/MSS.0000000000000556.
7. Rahemi H, Nigam N, Wakeling JM. The effect of intramuscular fat on skeletal muscle mechanics: implications for the elderly and obese. J R Soc Interface. 2015;12(109):20150365. 10.1098/rsif.2015.0365 . PubMed PMID: 26156300; PubMed Central PMCID: PMC4535407.
8. Ikezoe T, Mori N, Nakamura M, Ichihashi N. Age-related muscle atrophy in the lower extremities and daily physical activity in elderly women. Arch Gerontol Geriatr. 2011;53(2):e153–7. 10.1016/j.archger.2010.08.003.
9. Visser M, Kritchevsky SB, Goodpaster BH, Newman AB, Nevitt M, Stamm E, et al. Leg muscle mass and composition in relation to lower extremity performance in men and women aged 70 to 79: the health, aging and body composition study. J Am Geriatr Soc. 2002;50(5):897–904. 10.1046/j.1532-5415.2002.50217. .x PubMed PMID: 12028178.

10. Cawthon PM, Fox KM, Gandra SR, Delmonico MJ, Chiou CF, Anthony MS, et al. Do muscle mass, muscle density, strength, and physical function similarly influence risk of hospitalization in older adults? J Am Geriatr Soc. 2009;57(8):1411–9. 10.1111/j.1532-5415.2009.02366. .x PubMed PMID: 19682143; PubMed Central PMCID: PMC3269169.
11. Faulkner JA, Larkin LM, Claflin DR, Brooks SV. Age-related changes in the structure and function of skeletal muscles. Clin Exp Pharmacol Physiol. 2007;34(11):1091–6. 10.1111/j.1440-1681.2007. 04752.x PubMed PMID: 17880359.
12. Kalyani RR, Corriere M, Ferrucci L. Age-related and disease-related muscle loss: the effect of diabetes, obesity, and other diseases. Lancet Diabetes Endocrinol. 2014;2(10):819–29. 10.1016/S2213-8587(14)70034-8.
13. Bean JF, Kiely DK, Herman S, Leveille SG, Mizer K, Frontera WR, et al. The relationship between leg power and physical performance in mobility-limited older people. J Am Geriatr Soc. 2002;50(3):461–7. 10.1046/j.1532-5415.2002.50111. .x PubMed PMID: 11943041.
14. Callahan D, Phillips E, Carabello R, Frontera WR, Fielding RA. Assessment of lower extremity muscle power in functionally-limited elders. Aging Clin Exp Res. 2007;19(3):194–9. doi:10.1007/BF03324689 PubMed PMID: 17607086.
15. Casas-Herrero A, Cadore EL, Zambom-Ferraresi F, Idoate F, Millor N, Martínez-Ramirez A, et al. Functional capacity, muscle fat infiltration, power output, and cognitive impairment in institutionalized frail oldest old. Rejuvenation Res. 2013;16(5):396–403. 10.1089/rej.2013.1438 . PubMed PMID: 23822577; PubMed Central PMCID: PMC3804230.
16. Steves CJ, Mehta MM, Jackson SHD, Spector TD. Kicking Back Cognitive Ageing: Leg Power Predicts Cognitive Ageing after Ten Years in Older Female Twins. Gerontology. 2016;62(2):138–49. 10.1159/000441029.
17. Volpato S, Bianchi L, Lauretani F, Lauretani F, Bandinelli S, Guralnik JM, et al. Role of Muscle Mass and Muscle Quality in the Association Between Diabetes and Gait Speed. Diabetes Care. 2012;35(8):1672–9. 10.2337/dc11-2202.
18. de Lucena Alves CP, de Almeida SB, Lima DP, Neto PB, Miranda AL, Manini T, et al. Muscle Quality in Older Adults: A Scoping Review. J Am Med Dir Assoc. 2023;24(4):462–e46712. 10.1016/j.jamda.2023.02.012.
19. Santos R, Valamatos MJ, Mil-Homens P, Armada-da-Silva PAS. Muscle thickness and echo-intensity changes of the quadriceps femoris muscle during a strength training program. Radiography. 2018;24(4):e75–84. 10.1016/j.radi.2018.03.010.
20. Pillen S, van Alfen N. Skeletal muscle ultrasound. Neurol Res. 2011;33(10):1016–24. doi:10.1179/1743132811Y.0000000010 PubMed PMID: 22196753.
21. Watanabe Y, Yamada Y, Fukumoto, Yokoyama K, Yoshida T, Miyake, et al. Echo intensity obtained from ultrasonography images reflecting muscle strength in elderly men. Clin Interv Aging. 2013;993. 10.2147/CIA.S47263.

22. Takahashi Y, Fujino Y, Miura K, Toida A, Matsuda T, Makita S. Intra- and inter-rater reliability of rectus femoris muscle thickness measured using ultrasonography in healthy individuals. Ultrasound J. 2021;13(1):21. 10.1186/s13089-021-00224-8.
23. Filippo M, Lars AN, Maria S, Sandra AB. 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. 2019;5(3):037002. 10.1088/2057-1976/ab102f.
24. Hernández-Belmonte A, Martínez-Cava A, Pallarés JG. Panoramic ultrasound requires a trained operator and specific evaluation sites to maximize its sensitivity: A comprehensive analysis of the measurement errors. Physiol Behav. 2022;248:113737. 10.1016/j.physbeh.2022.113737.
25. Van Den Broeck J, Héréus S, Cattrysse E, Raeymaekers H, De Maeseneer M, Scafoglieri A. Reliability of Muscle Quantity and Quality Measured With Extended-Field-of-View Ultrasound at Nine Body Sites. Ultrasound Med Biol. 2023;49(7):1544–9. 10.1016/j.ultrasmedbio.2023.02.018.
26. Sarafraz H, Hadian MR, Ayoobi Yazdi N, Olyaei G, Bagheri H, Jalaie S, et al. Test-retest reliability of nerve and muscle morphometric characteristics utilizing ultrasound imaging in individuals with unilateral sciatica and controls. Chiropr Man Ther. 2018;26(1):47. 10.1186/s12998-018-0215-x.
27. Borg DN, Bach AJE, O'Brien JL, Sainani KL. Calculating sample size for reliability studies. PM&R. 2022;14(8):1018–25. 10.1002/pmrj.12850.
28. Pardo E, El Behi H, Boizeau P, Verdonk F, Alberti C, Lescot T. Reliability of ultrasound measurements of quadriceps muscle thickness in critically ill patients. BMC Anesthesiol. 2018;18(1):205. 10.1186/s12871-018-0647-9.
29. Bobak CA, Barr PJ, O'Malley AJ. Estimation of an inter-rater intra-class correlation coefficient that overcomes common assumption violations in the assessment of health measurement scales. BMC Med Res Methodol. 2018;18(1):93. 10.1186/s12874-018-0550-6.
30. Young HJ, Jenkins NT, Zhao Q, Mccully KK. Measurement of intramuscular fat by muscle echo intensity: Muscle Echo Intensity and Fat. Muscle Nerve. 2015;52(6):963–71. 10.1002/mus.24656.
31. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016;15(2):155–63. 10.1016/j.jcm.2016.02.012.
32. Sobolewski EJ, Wein LD, Crow JM, Carpenter KM. Intra-rater and inter-rater reliability of the process of obtaining cross-sectional area and echo intensity measurements of muscles from ultrasound images. J Ultrason. 2021;21(84):7–11. 10.15557/JoU.2021.0002 . PubMed PMID: 33791111; PubMed Central PMCID: PMC8008133.
33. Castro J, Livino De Carvalho K, Silva PE, Fachin-Martins E, Babault N, Marqueti RDC et al. Intra- and inter-rater reproducibility of ultrasound imaging of patellar and quadriceps tendons in critically ill patients. Patman S, editor. PLOS ONE. 2019;14(6):e0219057. 10.1371/journal.pone.0219057
34. Sions JM, Velasco TO, Teyhen DS, Hicks GE. Ultrasound imaging: intraexaminer and interexaminer reliability for multifidus muscle thickness assessment in adults aged 60 to 85 years versus younger adults. J Orthop Sports Phys Ther. 2014;44(6):425–34. 10.2519/jospt.2014.4584 . PubMed PMID: 24766356; PubMed Central PMCID: PMC4806552.

35. Pinto-Ramos J, Costa-Santos C, Costa F, Tavares H, Cabral J, Moreira T, et al. Reliability of point-of-care ultrasound for measuring quadriceps femoris muscle thickness. Eur J Phys Rehabil Med. 2022;58(5). 10.23736/S1973-9087.22.07432-9.
36. Takahashi Y, Fujino Y, Miura K, Toida A, Matsuda T, Makita S. Intra- and inter-rater reliability of rectus femoris muscle thickness measured using ultrasonography in healthy individuals. Ultrasound J. 2021;13(1):21. 10.1186/s13089-021-00224-8.
37. Ishida H, Suehiro T, Suzuki K, Yoneda T, Watanabe S. Influence of the ultrasound transducer tilt on muscle thickness and echo intensity of the rectus femoris muscle of healthy subjects. J Phys Ther Sci. 2017;29(12):2190–3. 10.1589/jpts.29.190.
38. AlMohimeed I, Ono Y. Ultrasound Measurement of Skeletal Muscle Contractile Parameters Using Flexible and Wearable Single-Element Ultrasonic Sensor. Sensors. 2020;20(13):3616. 10.3390/s20133616.

Figures

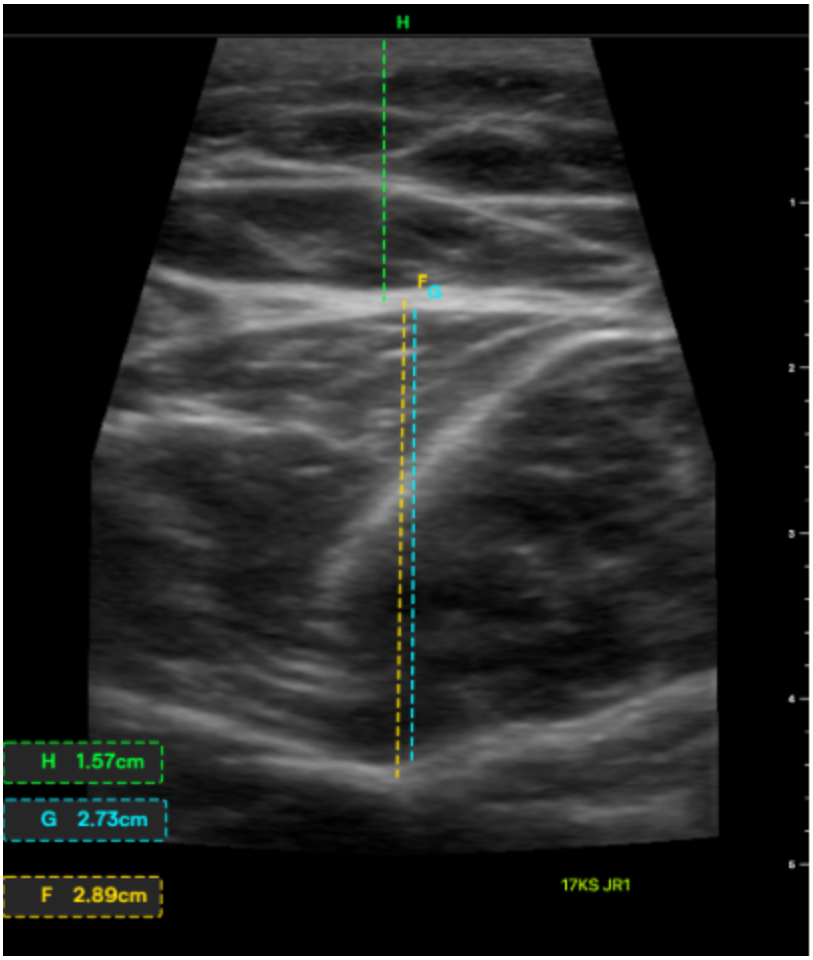

Figure 1

Example of landmarks used to measure muscle thickness adipose thickness.

Legend: MT, muscle thickness including the epimysium (F); MTex, muscle thickness excluding the epimysium (G); AT, adipose thickness (F).

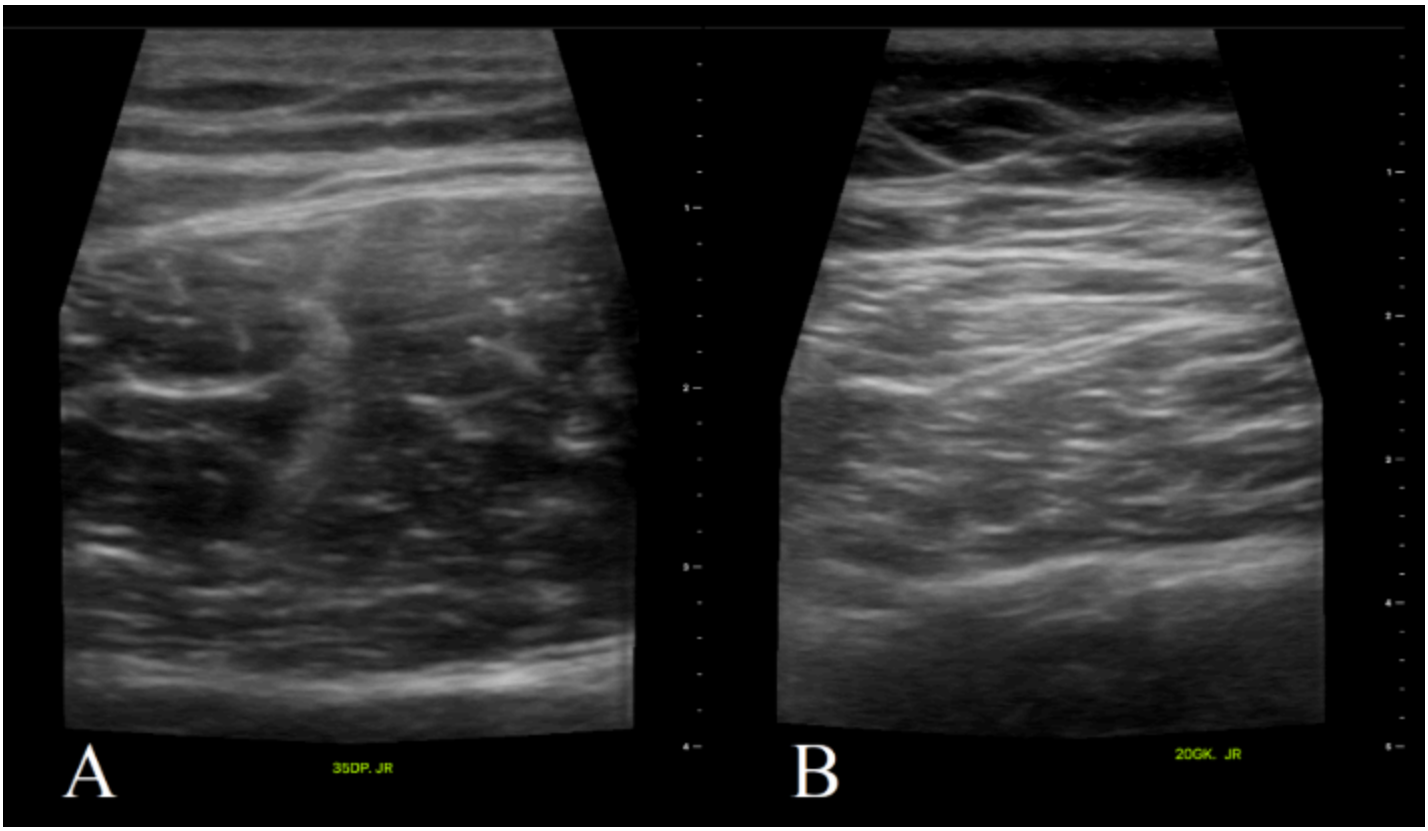

Figure 2

Example of an echo intensity measurement of two participants.

Legend: Panel A, participant with low echo intensity (EI); panel B, participant with high EI

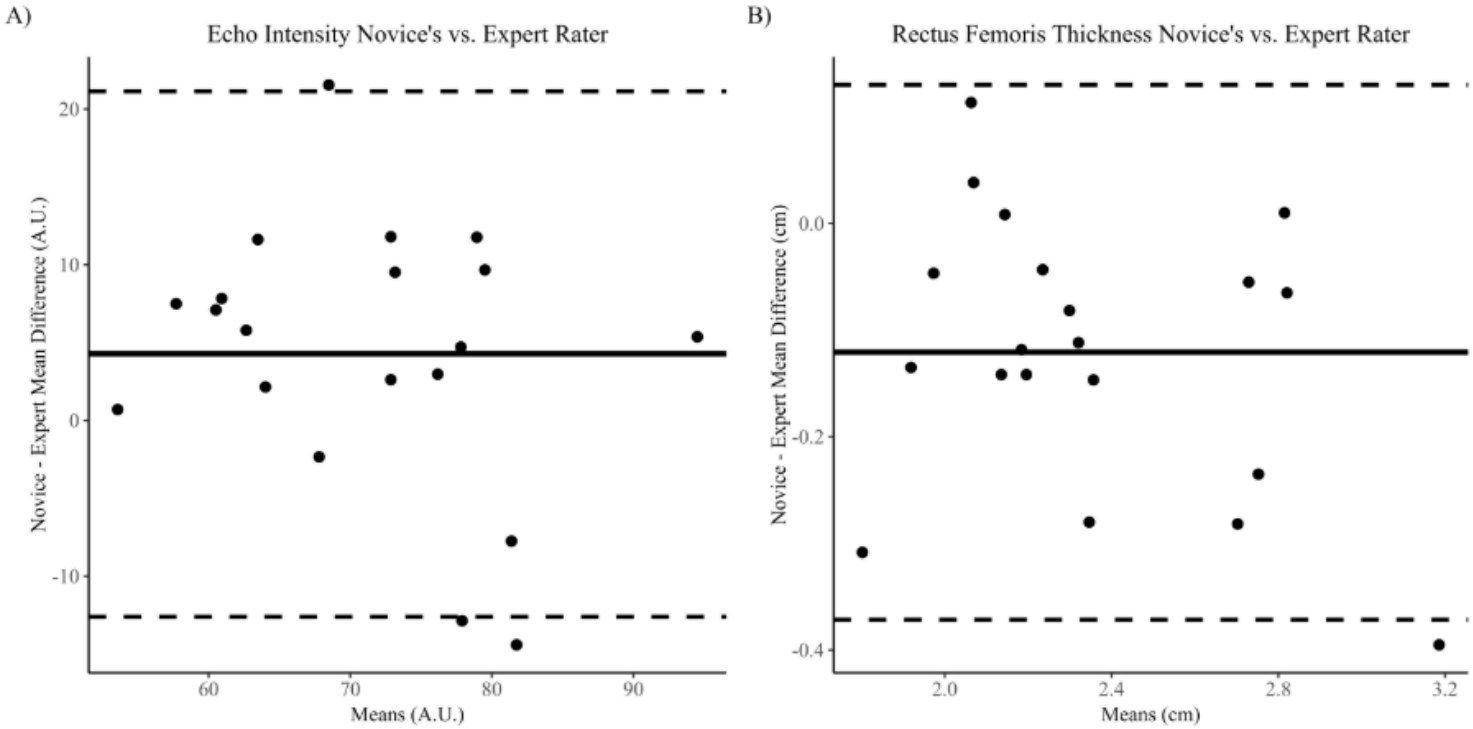

Figure 3

Bland–Altman plots of differences between novice and expert raters against the mean.

Legend: Panel A, echo intensity; panel B, rectus femoris thickness calculated without the epimysium. Solid horizontal lines represent the mean difference between the novice and expert raters. Dashed lines represent the limits of agreement. A.U. = arbitrary units; cm = centimetres.