



Article

# The Influence of Sonographer Experience on Skeletal Muscle Image Acquisition and Analysis

Joshua C. Carr <sup>1,2</sup>, Gena R. Gerstner <sup>3</sup>, Caleb C. Voskuil <sup>1</sup>, Joel E. Harden <sup>3</sup>, Dustin Dunnick <sup>4</sup>,  
Kristin M. Badillo <sup>5,6</sup>, Jason I. Pagan <sup>5,6</sup>, Kylie K. Harmon <sup>5,6</sup>, Ryan M. Girts <sup>5,6</sup>, Jonathan P. Beausejour <sup>5,6</sup>  
and Matt S. Stock <sup>5,6,\*</sup>

- <sup>1</sup> Neuromuscular Physiology Laboratory, Texas Christian University, Fort Worth, TX 76129, USA; joshua.carr@tcu.edu (J.C.C.); caleb.voskuil@tcu.edu (C.C.V.)  
<sup>2</sup> Department of Medical Education, TCU & UNTHSC School of Medicine, Fort Worth, TX 76129, USA  
<sup>3</sup> Neuromechanics Laboratory, Department of Human Movement Sciences, Old Dominion University, Norfolk, VA 23529, USA; ggerstne@odu.edu (G.R.G.); jhard042@odu.edu (J.E.H.)  
<sup>4</sup> Department of Health and Physical Education, Arkansas Tech University, Russellville, AR 72801, USA; ddunnick@atu.edu  
<sup>5</sup> Neuromuscular Plasticity Laboratory, Institute of Exercise Physiology and Rehabilitation Science, University of Central Florida, Orlando, FL 32816, USA; k.badillo22@knights.ucf.edu (K.M.B.); jason.pagan@ucf.edu (J.I.P.); kylie.harmon@ucf.edu (K.K.H.); ryan.girts@ucf.edu (R.M.G.); jonathan.beausejour@ucf.edu (J.P.B.)  
<sup>6</sup> School of Kinesiology and Physical Therapy, University of Central Florida, Orlando, FL 32816, USA  
\* Correspondence: matt.stock@ucf.edu



**Citation:** Carr, J.C.; Gerstner, G.R.; Voskuil, C.C.; Harden, J.E.; Dunnick, D.; Badillo, K.M.; Pagan, J.I.; Harmon, K.K.; Girts, R.M.; Beausejour, J.P.; et al. The Influence of Sonographer Experience on Skeletal Muscle Image Acquisition and Analysis. *J. Funct. Morphol. Kinesiol.* **2021**, *6*, 91. <https://doi.org/10.3390/jfmk6040091>

Academic Editor: Giuseppe Musumeci

Received: 8 October 2021  
Accepted: 25 October 2021  
Published: 29 October 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** The amount of experience with ultrasonography may influence measurement outcomes while images are acquired or analyzed. The purpose of this study was to identify the interrater reliability of ultrasound image acquisition and image analysis between experienced and novice sonographers and image analysts, respectively. Following a brief hands-on training session (2 h), the experienced and novice sonographers and analysts independently performed image acquisition and analyses on the biceps brachii, vastus lateralis, and medial gastrocnemius in a sample of healthy participants ( $n = 17$ ). Test–retest reliability statistics were computed for muscle thickness (transverse and sagittal planes), muscle cross-sectional area, echo intensity and subcutaneous adipose tissue thickness. The results show that image analysis experience generally has a greater impact on measurement outcomes than image acquisition experience. Interrater reliability for measurements of muscle size during image acquisition was generally good–excellent ( $ICC_{2,1}$ : 0.82–0.98), but poor–moderate for echo intensity ( $ICC_{2,1}$ : 0.43–0.77). For image analyses, interrater reliability for measurements of muscle size for the vastus lateralis and biceps brachii was poor–moderate ( $ICC_{2,1}$ : 0.48–0.70), but excellent for echo intensity ( $ICC_{2,1}$ : 0.90–0.98). Our findings have important implications for laboratories and clinics where members possess varying levels of ultrasound experience.

**Keywords:** ultrasonography; extended field of view; muscle size; muscle quality

## 1. Introduction

The use of ultrasonography for skeletal muscle imaging in the field of kinesiology is growing. This is likely a result of its affordability, validity, and reliability compared to advanced imaging techniques such as magnetic resonance imaging and computerized tomography [1–4]. Establishing the reliability of skeletal muscle ultrasound is critical since it is commonly used to assess skeletal muscle adaptations with exercise training [1,2,5], muscle disuse [1,6], aging [7–10] and disease [11,12]. Considerable evidence shows that ultrasound measurements demonstrate acceptable intra and interrater reliability [11,13–16]. However, the extent to which relative experience with ultrasound image acquisition and analyses influences its outcomes is not well defined.

A common scenario within a research laboratory or clinic is that its members possess varying levels of experience with a technique such as ultrasonography. This presents a challenge, as a critical aspect in longitudinal studies or patient evaluation relates to the feasibility of a single experimenter performing all ultrasound scans and analyses. The fact that relatively minor changes in probe orientation [17], pressure applied to the skin [18], or even scale calibration [16] during analysis can have marked effects on the outcomes illustrates the need to identify interrater reliability between vastly different ultrasound experience levels. Mayer et al. [11] have recently shown that following 8 h of expert-led ultrasound training, a small group of ultrasound-naïve physical therapy students had reliable ultrasound measurements compared to an expert sonographer. Importantly, the scans were performed on a group of patients within an intensive care unit, a group that had recovered from intensive care, and a healthy control, but the group sample sizes were small ( $n = 6$ ), several raters ( $n = 5$ ) of varying experience levels performed acquisition, and the influence of ultrasound analysis experience on the outcomes was not determined. Cleary et al. [19] recently show excellent interrater image analysis reliability for cross-sectional area, muscle thickness, and echo intensity in a group of novices when images were captured by the same, experienced sonographer. When examining interrater reliability for both image acquisition and analysis between novice and expert sonographers, Zaidman et al. [12] show similar outcomes between sonographers but these data were solely on the echo intensity values. Overall, the evidence suggests that ultrasound-derived measurements of skeletal muscle size and quality exhibit acceptable interrater reliability, yet there is insufficient data on how the relative experience of a sonographer influences both image acquisition and the analysis of muscle cross-sectional area, muscle thickness, subcutaneous adipose tissue thickness, and echo intensity.

The present experiment addresses how experience with ultrasonography influences image outcomes by identifying the interrater reliability for ultrasound image acquisition and the analyses between experienced and novice sonographers. An experienced and a novice sonographer performed image acquisition and the subsequent analyses were performed by an experienced and novice image analyst. Outcome measures consisted of muscle thickness, cross-sectional area, subcutaneous adipose tissue thickness, and echo intensity of three commonly studied muscles in the fields of kinesiology—the biceps brachii, vastus lateralis, and medial gastrocnemius muscles.

## 2. Materials and Methods

### 2.1. Study Design

A cross-sectional study design was used to examine the role of B-mode ultrasonography experience on the reliability of ultrasound-derived measurements of muscle thickness, cross-sectional area, echo intensity, and subcutaneous adipose tissue thickness. During a single visit to the University of Central Florida Institute of Exercise Physiology and Rehabilitation Science, participants underwent ultrasound imaging of the biceps brachii, vastus lateralis, and medial gastrocnemius muscles. An experienced and novice sonographer performed image acquisition and an experienced and novice image analyst performed image analysis. The order of testing between sonographers and muscles was randomized with a random number generator. Participants refrained from exercise for  $\geq 24$  h before their laboratory visit. All participants signed their Informed Consent, and this study was approved by the Institutional Review Board for Human Subjects at the University of Central Florida (IRB # STUDY00003175).

### 2.2. Participants

A total of 19 participants volunteered for this study and 17 were retained for analyses. The experienced and novice performed the scans together on the first two participants (one female and one male) as part of the hands-on training. Exclusion criteria were limited to neuromuscular or metabolic disease, a history of stroke, cancer, or heart attack, significant musculoskeletal pain, and use of medications that may impact physical performance.

Ten females (age =  $21 \pm 2$  yrs; stature =  $167.0 \pm 9.6$  cm; mass =  $61.3 \pm 8.3$  kg; ethnicity: Caucasian = 6, African American = 1, Hispanic = 3) and seven males (age =  $24 \pm 3$  years; stature =  $179.5 \pm 6.8$  cm; mass =  $81.9 \pm 7.6$  kg; ethnicity: Caucasian = 5, Hispanic = 2) were included in analysis.

### 2.3. Sonographers

An experienced (M.S.) and novice (J.C.) sonographer performed all ultrasound scans with the participants on a treatment table. The novice sonographer had never performed ultrasound measurements and was completely naïve to the methods, procedures, and requisite skills necessary to acquire ultrasound images. A brief custom-made video was crafted by an experienced sonographer (G.G.) on the research team regarding the basics of ultrasound image acquisition (Supplementary Materials). On the day of data collection, before acquisition, the experienced sonographer (M.S.) provided the novice with one-on-one instruction regarding the LOGIQ-E software interface, probe orientation, and scanning tips and pitfalls for the imaged muscles. Hands-on training was then accomplished by having the experienced (M.S.) and novice (J.C.) perform the ultrasound scans together on the first two participants. Following this, the experienced (M.S.) and novice (J.C.) performed all scans independently. In total, the novice (J.C.) had less than two hours of instruction before performing the scans without guidance or instruction. At the time of the experiment, the experienced sonographer (M.S.) had approximately seven years of experience with musculoskeletal sonography in adolescents, adults, and the elderly.

### 2.4. B-Mode Ultrasonography Image Acquisition

All images were taken from the right side of the participants while supine for the biceps brachii and vastus lateralis imaging and prone for the medial gastrocnemius assessment. The images were recorded with a portable B-mode imaging device (GE Logiq E BT12, GE Healthcare, Milwaukee, WI, USA) and a multi-frequency linear array probe (12 L-RS, 5–13 MHz, 38.4 mm field of view, GE Healthcare, Milwaukee, WI, USA) was used for the vastus lateralis, whereas a wideband linear array probe (L8-18i-RS, 4.5–18 MHz, 25 mm field of view, GE Healthcare, Milwaukee, WI, USA) was used for the biceps brachii and medial gastrocnemius muscles. All settings were kept consistent (Frequency 10 MHz, Gain 55 dB, Dynamic range 72, Depth 5 cm) across and within participants; however, a depth of 6 cm was required for three participants to view the full muscle and was kept constant across sonographers. Once the site was identified for the respective muscle, sharpie was applied to the skin surface before image acquisition and both sonographers used the identified site for probe placement. For each muscle, still images were captured in the sagittal and transverse planes, and then panoramic images were captured with the panoramic function (LogiqView, GE Healthcare, Milwaukee, WI, USA). Three images were captured for each scan for every muscle. For the panoramic images, the probe was oriented in the transverse plane and was guided by a flexible high-density foam pad to allow steady transverse movement of the probe across the imaging areas. For the biceps brachii, cloth tape was used to identify the 50% distance from the acromion process to the antecubital space. Similarly, the 50% distance between the greater trochanter and the superior border of the patella was used for the vastus lateralis. The site for the medial gastrocnemius muscle was determined on an individual-by-individual basis due to the large heterogeneity of the lower limb [20,21]. The site was identified by scanning in the transverse and sagittal planes of the muscle and visually identifying the site with the largest muscle thickness. A considerable amount of water-soluble transmission gel (Aquasonic 100 ultrasound transmission gel, Parker Laboratories, Inc., Fairfield, NJ, USA) was applied to the skin for all imaging.

### 2.5. Image Analysts

An experienced (G.G.) and novice (C.V.) image analyst were not present during data collection, therefore, were blind to the image coding and were unaware of who acquired the images. A brief custom-made video was crafted by an experienced sonographer (G.G.)

on the research team regarding the basics of ultrasound image analyses (Supplementary Materials). The experienced analyst (G.G.) instructed the novice on the procedures for the image analyses, addressing important steps and common challenges for the derived measurements. Hands-on training was then accomplished by having the experienced (G.G.) and novice (C.V.) perform the ultrasound analyses together on the first two participants. Following this, the experienced (G.G.) and novice (C.V.) performed all analyses independently. In total, the novice (C.V.) had less than two hours of instruction before analyzing the images without guidance or instruction. At the time of the experiment, the experienced analyzer (G.G.) had approximately six years of experience with musculoskeletal ultrasound image analyses in adolescents, adults, and the elderly.

### 2.6. B-Mode Ultrasonography Image Analysis

The ultrasound images were exported and analyzed with ImageJ software (version 1.53k; National Institutes of Health, Bethesda, MD, USA). The experienced analyst (G.G.) visually inspected the three images taken for each muscle and site (e.g., biceps brachii sagittal) and selected the clearest image of the three for analysis. The same selected image was then analyzed by the experienced (G.G.) and novice (C.V.) image analysts. The images were first scaled from pixels to cm using the straight-line function. Muscle thickness, in both sagittal and transverse plane, was quantified using the straight-line function at the midpoint of the muscle on the freeze-frame image. To quantify muscle cross-sectional area ( $\text{cm}^2$ ), the polygon function was used to outline the border of each muscle without any surrounding fascia on the panoramic image. Echo intensity was determined via gray-scale analysis using the histogram function within the same polygon used for cross-sectional area analyses. Using the same image that muscle cross-sectional area and echo intensity were outlined on, subcutaneous adipose tissue thickness was quantified using the straight-line function at three sites (medial, midpoint, lateral) from the skin to the superficial aponeurosis and calculated as the average of the three values.

### 2.7. Statistical Analysis

Descriptive statistics have been reported as the mean  $\pm$  SD for the following five variables: (1) muscle thickness in the sagittal plane, (2) muscle thickness in the transverse plane, (3) cross-sectional area, (4) echo intensity, and (5) subcutaneous adipose tissue thickness. Paired samples *t*-tests were performed to examine systematic variability, with an alpha level of 0.05 used to determine statistically significant differences. To provide insight into the precision and magnitude of the estimated differences, 95% confidence intervals (CIs) and Cohen's *d* effect sizes were computed, respectively. Cohen's *d* values of 0.2, 0.5, and 0.8 were used to classify small, moderate, and large differences, respectively [22]. The method of Bland and Altman [23] was used to identify the 95% limits of agreement between the experienced versus novice sonographers and image analysts. Reliability was quantified with intraclass correlation coefficients (ICCs) and computed with the 2-way random-effects model ( $\text{ICC}_{2,1}$ ) on account of its generalizability to other laboratories and testers [24,25]. The ICCs were evaluated based on a reliability scale where ICCs < 0.50 indicated "poor" reliability, ICCs of 0.50–0.75 indicated "moderate" reliability, ICCs of 0.75–0.90 indicated "good" reliability, and ICCs > 0.90 were indicative of "excellent" reliability [26]. The mean square error was used to calculate the absolute standard error of the measurement (SEM [expressed in absolute units and as a percentage of the grand mean]), and the minimal difference needed to be considered real (MD) statistics [25].

## 3. Results

### 3.1. Image Acquisition

Table 1 shows the mean  $\pm$  SD and % mean difference values for each of the five variables across the three different muscles. Generally, the images acquired by the novice sonographer resulted in larger thickness and cross-sectional area values, with mean differences ranging from 0.38 to 26.47%. Table 2 displays the reliability statistics. Statistically

significant differences ( $p < 0.05$ ) were observed for images acquired by the experienced versus novice sonographer for vastus lateralis echo intensity ( $p = 0.002$ ), medial gastrocnemius cross-sectional area ( $p = 0.035$ ), and biceps brachii subcutaneous tissue thickness ( $p = 0.002$ ), and these associated effect sizes were considered moderate or large. All non-significant differences were also associated with small or trivial effect sizes ( $d \leq 0.456$ ). Based on the ICCs, all variables were classified as demonstrating good–excellent reliability except medial gastrocnemius echo intensity (0.643) and biceps brachii echo intensity (0.437), and subcutaneous tissue thickness (0.740). Variables showing particularly poor SEMs included vastus lateralis subcutaneous tissue thickness (12.38%), biceps brachii sagittal thickness (11.97%), and subcutaneous tissue thickness (17.74%). Figure 1 shows example data, highlighting differences between sonographers.

**Table 1.** Mean ± SD values and the % mean difference between the experienced and novice sonographers for each muscle and variable.

	Transverse Thickness (cm)	Sagittal Thickness (cm)	CSA (cm <sup>2</sup> )	EI (A.U.)	Subcutaneous Thickness (cm)
<b>Vastus Lateralis</b>					
Experienced	2.39 ± 0.52	2.38 ± 0.54	27.04 ± 9.23	64.48 ± 8.96	1.17 ± 0.65
Novice	2.49 ± 0.41	2.46 ± 0.45	26.74 ± 8.65	59.32 ± 12.09	1.22 ± 0.62
% Mean Difference	4.18	3.36	1.11	8.00	4.27
<b>Medial Gastrocnemius</b>					
Experienced	1.86 ± 0.41	1.84 ± 0.36	11.74 ± 3.98	110.56 ± 9.84	0.585 ± 0.240
Novice	1.90 ± 0.38	1.90 ± 0.42	12.53 ± 4.93	110.14 ± 9.49	0.587 ± 0.248
% Mean Difference	2.15	3.26	6.73	0.38	1.72
<b>Biceps Brachii</b>					
Experienced	2.61 ± 0.77	2.44 ± 0.77	9.65 ± 5.54	120.87 ± 8.93	0.341 ± 0.144
Novice	2.79 ± 0.78	2.64 ± 0.92	10.00 ± 6.19	123.53 ± 11.87	0.431 ± 0.192
% Mean Difference	6.90	8.20	3.63	2.20	26.47

**Table 2.** Statistical outcomes for the image acquisition comparisons between the experienced versus novice sonographers. \* = statistically significant difference between sonographer images.

	Transverse Thickness (cm)	Sagittal Thickness (cm)	CSA (cm <sup>2</sup> )	EI (A.U.)	Subcutaneous Thickness (cm)
<b>Vastus Lateralis</b>					
<i>p</i>	0.136	0.288	0.417	0.002 *	0.334
<i>d</i>	0.381	0.267	0.202	0.898	0.241
95% CI	−0.237–0.035	−0.203–0.066	−0.468–1.074	2.208–8.123	−0.158–0.057
Limits of Agreement Upper–Lower Limit	−0.621–0.419	−0.594–0.452	−2.637–3.243	−6.110–16.440	−0.461–0.360
ICC <sub>2,1</sub>	0.826	0.854	0.986	0.77	0.945
SEM (raw units)	0.19	0.19	1.06	4.07	0.15
SEM (%)	7.69	7.79	3.95	6.57	12.38
MD	0.52	0.52	2.94	11.27	0.41
<b>Medial Gastrocnemius</b>					
<i>p</i>	0.226	0.105	0.035 *	0.838	0.92
<i>d</i>	0.305	0.416	0.559	0.05	0.025
95% CI	−0.118–0.030	−0.137–0.014	−1.512–−0.063	−3.859–4.697	−0.039–0.035
Limits of Agreement Upper–Lower Limit	−0.325–0.238	−0.351–0.228	−3.551–1.976	−15.888–16.727	−0.143–0.149
ICC <sub>2,1</sub>	0.932	0.92	0.939	0.643	0.959
SEM (raw units)	0.10	0.10	1.00	5.88	0.05
SEM (%)	5.41	5.58	8.22	5.33	8.71
MD	0.28	0.29	2.76	16.31	0.14

Table 2. Cont.

	Transverse Thickness (cm)	Sagittal Thickness (cm)	CSA (cm <sup>2</sup> )	EI (A.U.)	Subcutaneous Thickness (cm)
<b>Biceps Brachii</b>					
<i>p</i>	0.082	0.078	0.28	0.341	0.002 *
<i>d</i>	0.45	0.456	0.271	0.238	0.926
95% CI	−0.365–0.024	−0.418–0.025	−1.019–0.315	−8.394–3.078	−0.139–0.040
Limits of Agreement Upper–Lower Limit	−0.913–0.571	−1.040–0.647	−2.894–2.190	−24.524–19.207	−0.279–0.100
ICC <sub>2,1</sub>	0.867	0.856	0.975	0.437	0.74
SEM (raw units)	0.27	0.30	0.92	7.89	0.07
SEM (%)	9.91	11.97	9.34	6.46	17.74
MD	0.74	0.84	2.54	21.87	0.19

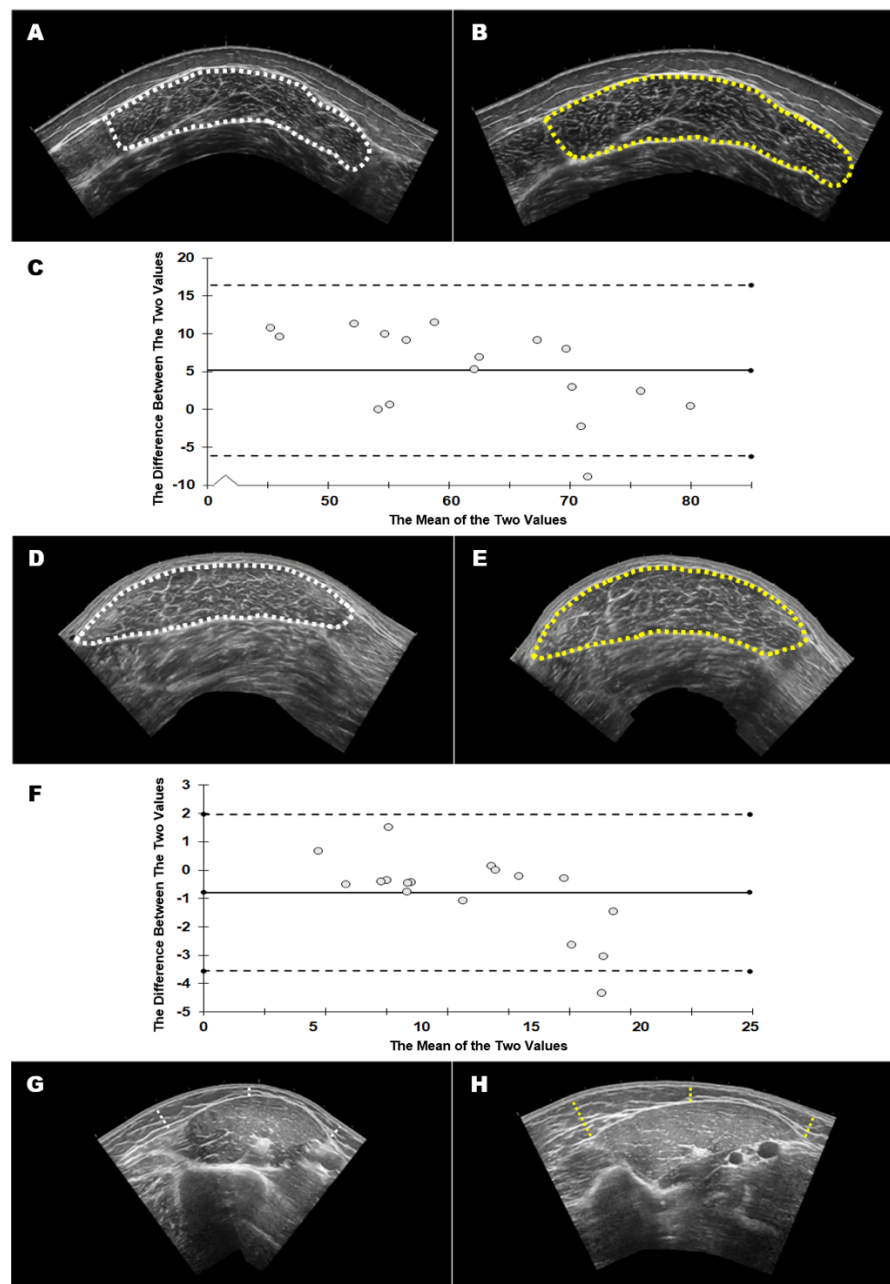
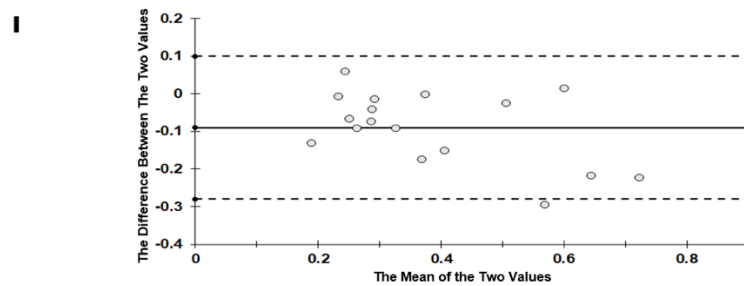


Figure 1. Cont.



**Figure 1.** Acquired images from the experienced sonographer (left column, **A,D,G**) and the novice sonographer (right column, **B,E,H**) were analyzed by the experienced analyst (G.G.). The variables that were significantly different between sonographers during acquisition are shown: echo intensity of the vastus lateralis (top row, **A,B**), cross sectional area of the medial gastrocnemius (middle row, **D,E**), and subcutaneous adipose tissue thickness of the biceps brachii (bottom row, **G,H**). Bland Altman plots are shown for each variable (**C,F,I**). Dotted lines: 95% limits of agreement.

### 3.2. Image Analysis

Table 3 shows the mean ± SD and % mean difference values for each of the five variables across the three different muscles. Qualitatively, there were no consistent patterns of lower or greater values being demonstrated across variables or muscles. The mean differences ranged from 0.04 to 14.11%. Table 4 displays the reliability statistics. Statistically significant differences ( $p < 0.05$ ) were observed for images analyzed by the experienced versus novice investigator for vastus lateralis cross-sectional area ( $p = 0.005$ ) and biceps brachii transverse thickness ( $p = 0.029$ ). Only vastus lateralis cross-sectional area demonstrated an effect size that was considered moderate ( $d = 0.514$ ). All other effect sizes were considered small ( $d \leq 0.393$ ). Based on the ICCs, five of the variables demonstrated moderate reliability and one variable showed poor reliability. Eight out of the 15 variables demonstrated SEMs  $\geq 10.0\%$ . Figure 2 shows example data, highlighting differences between image analysts.

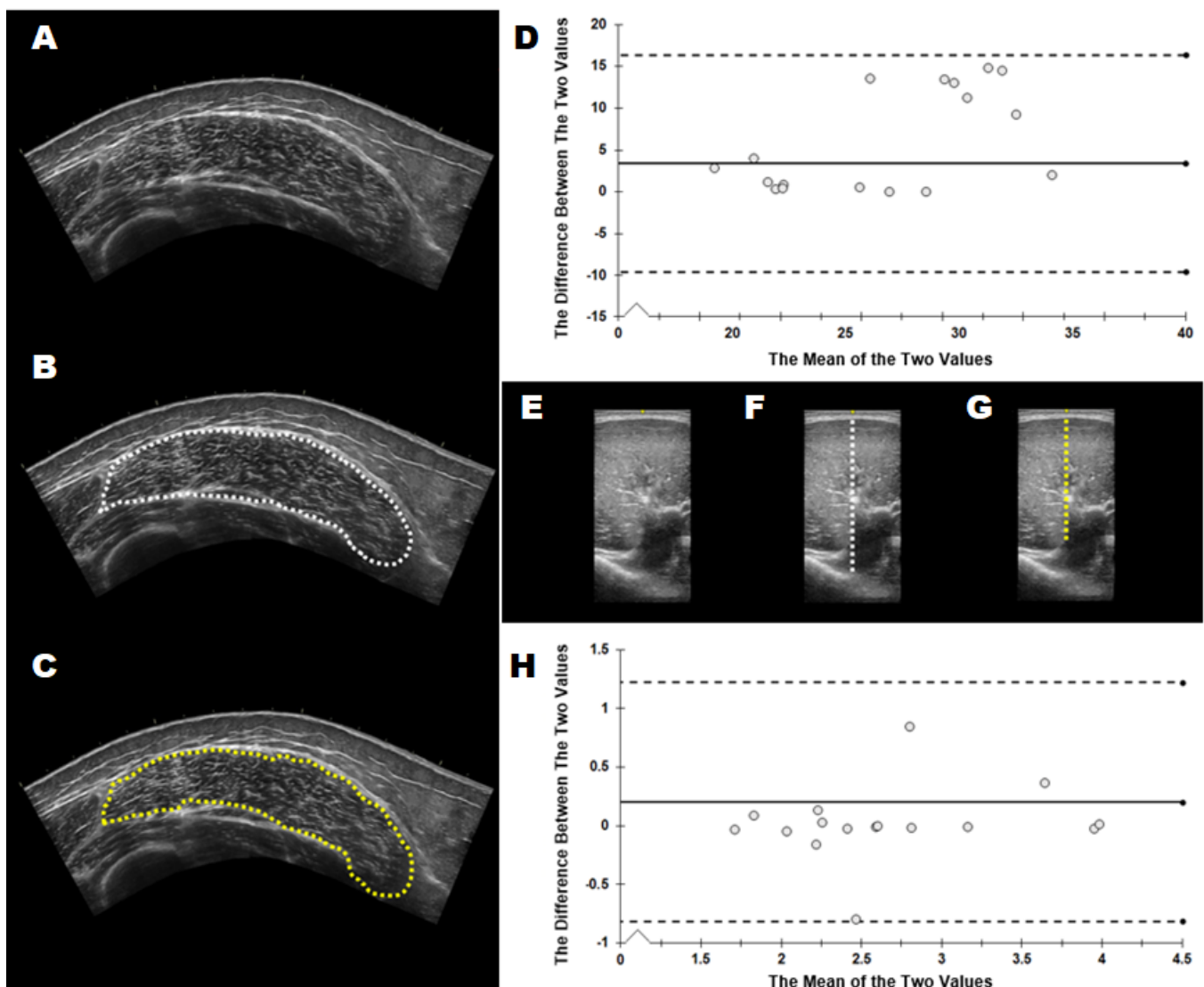
**Table 3.** Mean ± SD values and the % mean difference between the experienced and novice image analysts for each muscle and variable.

	Transverse Thickness (cm)	Sagittal Thickness (cm)	CSA (cm <sup>2</sup> )	EI (A.U.)	Subcutaneous Thickness (cm)
<b>Vastus Lateralis</b>					
<b>Experienced</b>	2.44 ± 0.46	2.42 ± 0.49	26.89 ± 8.81	61.90 ± 10.80	1.20 ± 0.62
<b>Novice</b>	2.58 ± 0.54	2.58 ± 0.54	23.49 ± 8.12	61.98 ± 11.52	1.12 ± 0.52
<b>% Mean Difference</b>	5.95	6.78	12.64	0.13	6.40
<b>Medial Gastrocnemius</b>					
<b>Experienced</b>	1.87 ± 0.39	1.87 ± 0.38	12.13 ± 4.43	110.35 ± 9.52	0.586 ± 0.241
<b>Novice</b>	1.91 ± 0.43	1.92 ± 0.45	11.67 ± 4.36	110.84 ± 9.46	0.568 ± 0.254
<b>% Mean Difference</b>	1.47	2.47	3.84	0.44	3.10
<b>Biceps Brachii</b>					
<b>Experienced</b>	2.70 ± 0.77	2.54 ± 0.84	9.82 ± 5.78	122.20 ± 10.43	0.386 ± 0.173
<b>Novice</b>	2.50 ± 0.64	2.55 ± 0.75	8.44 ± 3.94	122.25 ± 10.84	0.399 ± 0.210
<b>% Mean Difference</b>	7.55	0.20	14.11	0.04	3.45

**Table 4.** Statistical outcomes for the image analysis comparisons between the experienced versus novice analysts.  
 \* = statistically significant difference between image analysts.

	<b>Transverse Thickness (cm)</b>	<b>Sagittal Thickness (cm)</b>	<b>CSA (cm<sup>2</sup>)</b>	<b>EI (A.U.)</b>	<b>Subcutaneous Thickness (cm)</b>
<b>Vastus Lateralis</b>					
<i>p</i>	0.087	0.07	0.005 *	0.827	0.222
<i>d</i>	0.303	0.321	0.514	0.038	0.214
95% CI	−0.312 to 0.022	−0.342 to 0.014	1.090 to 5.709	−0.807 to 0.650	−0.049 to 0.202
Limits of Agreement	−1.084 to 0.794	−1.165 to 0.837	−9.572 to 16.371	−4.171 to 4.014	−0.626 to 0.780
Upper–Lower Limit					
ICC <sub>2,1</sub>	0.527	0.489	0.648	0.983	0.801
SEM (raw units)	0.34	0.36	4.68	1.48	0.25
SEM (%)	13.5	14.4	18.6	2.4	21.9
MD	0.94	1.00	12.97	4.09	0.7
<b>Medial Gastrocnemius</b>					
<i>p</i>	0.313	0.178	0.093	0.307	0.158
<i>d</i>	0.176	0.236	0.296	0.178	0.248
95% CI	−0.083 to 0.027	−0.115 to 0.022	−0.082 to 1.015	−1.452 to 0.471	−0.007 to 0.044
Limits of Agreement	−0.337 to 0.281	−0.430 to 0.338	−2.615 to 3.547	−5.890 to 4.910	−0.125 to 0.162
Upper–Lower Limit					
ICC <sub>2,1</sub>	0.927	0.887	0.933	0.958	0.955
SEM (raw units)	0.11	0.14	1.11	1.95	0.05
SEM (%)	5.9	7.3	9.3	1.8	9.0
MD	0.31	0.38	3.08	5.4	0.14
<b>Biceps Brachii</b>					
<i>p</i>	0.029 *	0.952	0.052	0.953	0.486
<i>d</i>	0.393	0.01	0.346	0.01	0.121
95% CI	0.023 to 0.385	−0.175 to 0.165	−0.012 to 2.783	−1.675 to 1.579	−0.052 to 0.025
Limits of Agreement	−0.815 to 1.223	−0.960 to 0.950	−6.465 to 9.236	−9.190 to 9.094	−0.229 to 0.202
Upper–Lower Limit					
ICC <sub>2,1</sub>	0.709	0.818	0.653	0.906	0.839
SEM (raw units)	0.37	0.35	2.83	3.3	0.08
SEM (%)	14.1	13.5	31	2.7	19.8
MD	1.02	0.96	7.85	9.14	0.22





**Figure 2.** Acquired images (A,E) from the experienced sonographer (M.S.) were analyzed by the experienced analyst (B,F) and the novice analyst (C,G). Representative images for the variables that were significantly different between analysts during image analysis are shown: cross sectional area of the vastus lateralis (B,C) and transverse thickness of the biceps brachii (F,G). Bland Altman plots are shown for vastus lateralis cross-sectional area (D) and biceps brachii thickness (H). Dotted lines: 95% limits of agreement.

#### 4. Discussion

This experiment describes how the relative experience with ultrasound image acquisition and analyses influences outcomes of skeletal muscle size and quality. The findings from this study show that experience with image acquisition and analysis generally has small effects ( $d < 0.30$ ) and good–excellent ( $ICC_{2,1} > 0.80$ – $0.98$ ) interrater reliability on most ultrasound outcomes. However, significant differences and large effects between experienced and novice sonographers were observed for image acquisition and analysis for some of the variables. For image acquisition, medial gastrocnemius cross-sectional area, vastus lateralis echo intensity, and biceps brachii subcutaneous adipose tissue thickness were significantly different between sonographers. Despite this, experienced–novice interrater reliability for measures of muscle size was good–excellent ( $ICC_{2,1} > 0.82$ – $0.98$ ). Similarly, for analysis, vastus lateralis cross-sectional area and biceps brachii muscle thickness were significantly different between sonographers, and measures of muscle size for the vastus lateralis exhibited poor reliability ( $ICC_{2,1} < 0.65$ ), yet interrater reliability for echo intensity

was excellent ( $ICC_{2,1} > 0.90$ ) for all muscles. The present data show that relative experience with ultrasound techniques has a task-specific influence on image outcomes that should be considered when designing and interpreting ultrasound-based assessments.

There have been limited attempts to quantify the influence of ultrasound experience on image outcomes [11,12,19]. The rationale for this comparison is based on the fact that a laboratory or clinic has a continual rotation of proficiency levels across time. Identifying reliability between high and low experience levels is necessary for study design and coordination. The present findings generally agree with similar reliability studies that have compared individuals with extensive versus limited ultrasound experience [11,12,19]; however, we report greater systematic variability between experienced versus novice sonographers than previously shown [11,12]. The strengths of the present data are that we show how ultrasound acquisition and analysis experience influences the outcomes for the variables used to determine muscle size and quality—muscle cross-sectional area, muscle thickness, subcutaneous adipose thickness, and echo intensity. The poor interrater reliability for echo intensity during image acquisition is likely an artifact of differences in angle, placement, and possibly the speed of the ultrasound probe during acquisition. Whereas the poor interrater reliability for vastus lateralis and biceps brachii muscle size analyses is likely explained by the inability of the novice to discern fascial borders due to their limited experience and challenging shapes of these muscles. As such, the level of skill that is required to accurately acquire ultrasound images may be muscle and variable dependent.

#### 4.1. Muscle Cross-Sectional Area

Ultrasound-derived measurements of muscle cross-sectional area have been cross-validated against MRI and CT imaging and show good agreements [2–4]. It has been suggested that sonographer proficiency is needed when collecting panoramic images with the extended field of view technique to acquire high-quality images [2]. Indeed the degree of muscle curvature can present challenges and data show that reliability weakens at regions with greater relative curvature, such as the distal portion of the thigh [2,3]. The present data support these observations as the medial gastrocnemius shows greater systematic variability compared to the biceps brachii and vastus lateralis. Nevertheless, excellent interrater reliability is shown for the vastus lateralis ( $ICCs > 0.90$ ) [2–4], the medial gastrocnemius ( $ICCs > 0.90$ ) [5], and the biceps brachii ( $ICCs > 0.90$ ) [11] muscles. Our findings extend these observations by showing that a novice sonographer can acquire reliable extended field of view images in these muscles, but experienced–novice disparity increases with technical demand likely due to greater anatomical contour. Interestingly, the experienced–novice comparison for image acquisition versus image analysis shows that reliability was substantially weaker for image analysis. This is an important finding because it demonstrates that extended field of view image analyses requires sonographer proficiency in addition to the skills required for high-quality image acquisition. The present data show that the minimal differences for image acquisition ( $MD = 2.94, 2.54, 2.76 \text{ cm}^2$ ) were smaller than those of analysis ( $MD = 12.97, 7.85, 3.08 \text{ cm}^2$ ) for the vastus lateralis, biceps brachii, and medial gastrocnemius, respectively. There are three critical points to consider based on these data. One, the SEM for experienced–novice image acquisition ( $0.92\text{--}1.06 \text{ cm}^2$ ) is similar to that shown for the SEM of ultrasound compared to MRI ( $0.87 \text{ cm}^2$ ) [4] and CT ( $0.1\text{--}1.1 \text{ cm}^2$ ) [2]. Two, the minimal differences for experienced–novice image acquisition are likely small enough to detect resistance training-induced muscle hypertrophy and disuse induced atrophy for the lower limb [1,2,4]. Finally, three, the low level of interrater reliability for image analysis would have been unable to detect those effects. Similar outcomes have recently been shown for image analysis interrater reliability in novice sonographers [19]. Despite excellent ICC values, systematic variability was evident for measures of muscle size for the vastus lateralis, rectus femoris, and first dorsal interosseus among three different novice sonographers [19]. Collectively, it seems that experience level affects extended field of view image analysis more than image acquisition.

#### 4.2. Echo Intensity and Subcutaneous Adipose Tissue Thickness

The grayscale analysis that determines skeletal muscle echo intensity values is affected by relatively minor alterations in probe positioning mechanics [17]. Given this, it is unsurprising that interrater reliability for echo intensity during acquisition was generally poor–moderate and weaker than the interrater reliability values for image analysis. Zaidman et al. [12] suggest that minimal training is necessary to acquire reliable and clinically valid measures of echo intensity. The authors show that following only a 20-min expert-led training session, interrater reliability for echo intensity between a novice and expert was highly reliable ( $ICC \geq 0.85$ ). It is important to point out that echo intensity was derived from polygon tracing of panoramic images in the present study [19], not the rectangle function from still images [11,12]. This distinction likely explains the differing levels of interrater reliability for image acquisition between experience levels in the present study and others [11,12]. The interrater reliability values of the present study for image analysis are similar to those reported for three novice sonographers [19]. The experienced–novice difference in subcutaneous adipose tissue thickness for the biceps brachii during image acquisition is additional evidence that probe mechanics differed between sonographers, emphasizing the importance of probe pressure and alignment during image acquisition. Although the reliability statistics for echo intensity during image acquisition were weaker than those for analysis, the values for the vastus lateralis are comparable to test–retest reliability measures performed by the same, experienced sonographer [6]. Nevertheless, combining the observations of experienced–novice differences in subcutaneous thickness with the comparatively higher SEM and MD values for echo intensity during image acquisition versus analysis, it seems that a novice sonographer requires more than minimal training to acquire, but not analyze, reliable measures of echo intensity.

#### 4.3. Muscle Thickness

Jenkins et al. [14] suggest that given the greater skill required for extended field of view imaging, transverse imaging may be a more convenient option for measurements of muscle size. In support, the authors show excellent reliability for cross-sectional area, muscle thickness, and echo intensity determined from panoramic and a single transverse image for the biceps brachii with strong association between muscle cross-sectional area and muscle thickness ( $r = 0.93$ ). The interrater image acquisition reliability values for muscle thickness obtained in the transverse and sagittal planes show good–excellent reliability with minimal difference values sensitive enough to detect resistance training-induced increases in muscle thickness following longer (>6 weeks) training interventions [5,27,28], but likely not short-term training induced hypertrophy [29]. The SEM for muscle thickness measurements during acquisition are similar to those shown by a single experienced sonographer [30] and those by Mayer et al. [11] with raters of different experience level. Like the issues encountered with muscle cross-sectional area, the interrater reliability for image analysis was considerably weaker for the vastus lateralis and biceps brachii muscles compared to image acquisition. This was not the case for the medial gastrocnemius muscle, likely due to the brightness and clarity of the fascial borders for this muscle compared to the others in this study. Figure 2 shows how the image analysts differed in their muscle size measurements for vastus lateralis cross-sectional area and biceps brachii muscle thickness. The inability of the novice image analyst to consistently identify muscle boundaries and trace faint fascial borders is likely a major factor contributing to the poor interrater reliability for muscle size during image analyses.

#### 4.4. Ultrasound Considerations for Novice Onboarding

We recommend that laboratories use structured onboarding procedures for the novice sonographer. We recommend the following considerations for the novice sonographer during onboarding.

Laboratory standards: a video demonstrating image acquisition and analysis procedures for the specific equipment provides an accessible and convenient means to stan-

standardize laboratory procedures, techniques, and instructions for the novice (Supplementary Materials). This should be followed up with hands-on experiential learning practices on different skeletal muscles and persons with an experienced sonographer.

**Demonstrations:** how probe mechanics [16–18] influence the ultrasound image should be emphasized during acquisition training. Representative images demonstrating the fascial borders should be provided to the novice during analysis training for the skeletal muscles of interest and their surrounding anatomy, particularly challenging images with faint fascial borders. Since extended field of view scanning and analyses are more technically demanding, these scans should be emphasized and integrated.

**Time:** the total training time for novice onboarding is challenging to recommend. The formal training time for the present study was ~2 h, whereas ~8 h of training was performed by Mayer et al. [8] who show greater interrater reliability for image acquisition in a more heterogeneous sample. Regardless, recording the amount of time spent training for the novice provides an objective way to monitor their experience. To this point, keeping a formal time log for everyone may be a convenient method to quantify experience levels within a laboratory or clinic.

#### 4.5. Limitations

The present data have some important limitations to consider. First, although the sonographers performed the scans independently, the scans were performed on the same site that was determined (and marked) by the first sonographer. It may be that individuals of differing experience would acquire images from different relative muscle locations, yet this was not entirely the question we were attempting to answer. Another limitation is that we did not capture and compare other ultrasound-derived measurements such as fascicle length and pennation angle between novice and experienced sonographers. However, with the rise in automated analysis methods for these measurements, comparisons between an experienced sonographer and the automated analysis are needed [31]. Lastly, although the present sample of participants varied in their anthropometry, training status, and ethnicity, they were all young adults (<27 years) free of disease, illness, and injury. Future studies are encouraged to identify interrater reliability across a longitudinal intervention (i.e., resistance training, disuse, aging) to describe whether different sonographers capture the same magnitude of the respective effect.

#### 5. Conclusions

The present data show how experience with ultrasound image acquisition and analysis influences measurements of skeletal muscle size and quality. Since ultrasound imaging is a relatively simple procedure, it offers a lower barrier to entry for skeletal muscle measurements compared to other techniques. This has spurred interest in the question, how much training is required for ultrasound proficiency? Despite the level of convenience for ultrasound image acquisition and analysis, our data show that a tradeoff exists for interrater reliability between experienced and novice sonographers during image acquisition versus analyses. The experienced–novice comparisons for image acquisition show measures of muscle size can be reliably acquired, but measurements of muscle quality cannot. Whereas experienced–novice comparisons for image analyses indicate that measures of muscle quality are reliably analyzed, but measurements of muscle size for the vastus lateralis and biceps brachii are not. Many authors have shown that skill is required for high-quality ultrasound image acquisition, yet a critical interpretation from the present study is that ultrasound image analyses are not trivial procedures. Comparatively speaking, the experienced–novice differences were more severe for image analyses than image acquisition. Collectively, these findings suggest that ultrasound image analysis experience has a greater influence on the derived outcome variables than acquisition experience. The findings of this study have implications for laboratories that use ultrasonography and possess members of varying experience levels.

**Supplementary Materials:** The educational videos created by G.G. for the novice sonographers can be found at: [https://www.youtube.com/watch?v=7Dsyr\\_JwkaU](https://www.youtube.com/watch?v=7Dsyr_JwkaU); <https://www.youtube.com/watch?v=FgV5evU9fTE>.

**Author Contributions:** Conceptualization, M.S.S., J.C.C. and G.R.G.; methodology, M.S.S., J.C.C. and G.R.G.; software, G.R.G. and C.C.V.; formal analysis, M.S.S., J.C.C., G.R.G., C.C.V. and J.P.B.; investigation, All authors; resources, M.S.S. and G.R.G.; writing—original draft preparation, M.S.S., J.C.C. and G.R.G.; writing—review and editing, M.S.S., J.C.C. and G.R.G.; supervision, M.S.S., J.C.C. and G.R.G.; project administration, M.S.S. and K.M.B.; investigation, J.E.H., D.D., J.I.P., K.K.H. and R.M.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** This study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board of the University of Central Florida (IRB # STUDY00003175, date of approval: 30 June 2021).

**Informed Consent Statement:** Informed consent was obtained from all participants involved in this study.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author. The data are not publicly available.

**Acknowledgments:** The authors would like to thank the research participants for their willingness to take part in this study.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Stokes, T.; Tripp, T.R.; Murphy, K.; Morton, R.W.; Oikawa, S.Y.; Choi, H.L.; McGrath, J.; McGlory, C.; MacDonald, M.J.; Phillips, S.M. Methodological Considerations for and Validation of the Ultrasonographic Determination of Human Skeletal Muscle Hypertrophy and Atrophy. *Physiol. Rep.* **2021**, *9*, e14683. [[CrossRef](#)]
2. Noorkoiv, M.; Nosaka, K.; Blazevich, A.J. Assessment of Quadriceps Muscle Cross-Sectional Area by Ultrasound Extended-Field-of-View Imaging. *Eur. J. Appl. Physiol.* **2010**, *109*, 631–639. [[CrossRef](#)] [[PubMed](#)]
3. Scott, J.M.; Martin, D.S.; Ploutz-Snyder, R.; Caine, T.; Matz, T.; Arzeno, N.M.; Buxton, R.; Ploutz-Snyder, L. Reliability and Validity of Panoramic Ultrasound for Muscle Quantification. *Ultrasound Med. Biol.* **2012**, *38*, 1656–1661. [[CrossRef](#)]
4. Ahtiainen, J.P.; Hoffren, M.; Hulmi, J.J.; Pietikäinen, M.; Mero, A.A.; Avela, J.; Häkkinen, K. Panoramic Ultrasonography Is a Valid Method to Measure Changes in Skeletal Muscle Cross-Sectional Area. *Eur. J. Appl. Physiol.* **2009**, *108*, 273. [[CrossRef](#)]
5. Abe, T.; DeHoyos, D.V.; Pollock, M.L.; Garzarella, L. Time Course for Strength and Muscle Thickness Changes Following Upper and Lower Body Resistance Training in Men and Women. *Eur. J. Appl. Physiol.* **2000**, *81*, 174–180. [[CrossRef](#)]
6. MacLennan, R.J.; Sahebi, M.; Becker, N.; Davis, E.; Garcia, J.M.; Stock, M.S. Declines in Skeletal Muscle Quality vs. Size Following Two Weeks of Knee Joint Immobilization. *PeerJ* **2020**, *8*, e8224. [[CrossRef](#)] [[PubMed](#)]
7. Stock, M.S.; Whitson, M.; Burton, A.M.; Dawson, N.T.; Sobolewski, E.J.; Thompson, B.J. Echo Intensity Versus Muscle Function Correlations in Older Adults Are Influenced by Subcutaneous Fat Thickness. *Ultrasound Med. Biol.* **2018**, *44*, 1597–1605. [[CrossRef](#)] [[PubMed](#)]
8. Stock, M.S.; Mota, J.A.; Hernandez, J.M.; Thompson, B.J. Echo Intensity and Muscle Thickness as Predictors of Athleticism and Isometric Strength in Middle-School Boys. *Muscle Nerve* **2017**, *55*, 685–692. [[CrossRef](#)]
9. Gerstner, G.; Thompson, B.; Rosenberg, J.; Sobolewski, E.; Scharville, M.; Ryan, E. Neural and Muscular Contributions to the Age-Related Reductions in Rapid Strength. *Med. Sci. Sports Exerc.* **2017**, *49*, 1331–1339. [[CrossRef](#)]
10. Gerstner, G.R.; Giuliani, H.K.; Mota, J.A.; Ryan, E.D. Age-Related Reductions in Muscle Quality Influence the Relative Differences in Strength and Power. *Exp. Gerontol.* **2017**, *99*, 27–34. [[CrossRef](#)]
11. Mayer, K.P.; Dhar, S.; Cassity, E.; Denham, A.; England, J.; Morris, P.E.; Dupont-Versteegden, E.E. Interrater Reliability of Muscle Ultrasonography Image Acquisition by Physical Therapists in Patients Who Have or Who Survived Critical Illness. *Phys. Ther.* **2020**, *100*, 1701–1711. [[CrossRef](#)]
12. Zaidman, C.M.; Wu, J.S.; Wilder, S.; Darras, B.T.; Rutkove, S.B. Minimal Training Is Required to Reliably Perform Quantitative Ultrasound of Muscle. *Muscle Nerve* **2014**, *50*, 124–128. [[CrossRef](#)]
13. Rosenberg, J.G.; Ryan, E.D.; Sobolewski, E.J.; Scharville, M.J.; Thompson, B.J.; King, G.E. Reliability of Panoramic Ultrasound Imaging to Simultaneously Examine Muscle Size and Quality of the Medial Gastrocnemius. *Muscle Nerve* **2014**, *49*, 736–740. [[CrossRef](#)]
14. Jenkins, N.D.M.; Miller, J.M.; Buckner, S.L.; Cochrane, K.C.; Bergstrom, H.C.; Hill, E.C.; Smith, C.M.; Housh, T.J.; Cramer, J.T. 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. [[CrossRef](#)]

15. Tomko, P.M.; Muddle, T.W.; Magrini, M.A.; Colquhoun, R.J.; Luera, M.J.; Jenkins, N.D. Reliability and Differences in Quadriceps Femoris Muscle Morphology Using Ultrasonography: The Effects of Body Position and Rest Time. *Ultrasound* **2018**, *26*, 214–221. [[CrossRef](#)] [[PubMed](#)]
16. Sobolewski, E.J.; Wein, L.D.; Crow, J.M.; Carpenter, K.M. 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*, 7–11. [[CrossRef](#)] [[PubMed](#)]
17. Dankel, S.J.; Abe, T.; Bell, Z.W.; Jessee, M.B.; Buckner, S.L.; Mattocks, K.T.; Mouser, J.G.; Loenneke, J.P. The Impact of Ultrasound Probe Tilt on Muscle Thickness and Echo-Intensity: A Cross-Sectional Study. *J. Clin. Densitom.* **2020**, *23*, 630–638. [[CrossRef](#)]
18. Treece, G.M.; Prager, R.W.; Gee, A.H.; Berman, L. Correction of Probe Pressure Artifacts in Freehand 3D Ultrasound. *Med. Image Anal.* **2002**, *6*, 199–214. [[CrossRef](#)]
19. Cleary, C.; Nabavizadeh, O.; Young, K.; Herda, A. *Inter- and Intra-Rater Reliability of Ultrasound-Derived Muscle Quantification*, 2021.
20. Wang, J.; Hu, Y.; Tian, G. Ultrasound Measurements of Gastrocnemius Muscle Thickness in Older People with Sarcopenia. *Clin. Interv. Aging* **2018**, *13*, 2193–2199. [[CrossRef](#)]
21. Barber, L.; Barrett, R.; Lichtwark, G. Validity and Reliability of a Simple Ultrasound Approach to Measure Medial Gastrocnemius Muscle Length. *J. Anat.* **2011**, *218*, 637–642. [[CrossRef](#)]
22. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed.; L. Erlbaum Associates: Hillsdale, MI, USA, 1988; ISBN 978-0-8058-0283-2.
23. Bland, J.M.; Altman, D.G. Statistical Methods for Assessing Agreement between Two Methods of Clinical Measurement. *Lancet Lond. Engl.* **1986**, *1*, 307–310. [[CrossRef](#)]
24. Shrout, P.E.; Fleiss, J.L. Intraclass Correlations: Uses in Assessing Rater Reliability. *Psychol. Bull.* **1979**, *86*, 420–428. [[CrossRef](#)] [[PubMed](#)]
25. Weir, J.P. Quantifying Test-Retest Reliability Using the Intraclass Correlation Coefficient and the SEM. *J. Strength Cond. Res.* **2005**, *19*, 231. [[CrossRef](#)] [[PubMed](#)]
26. Koo, T.K.; Li, M.Y. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J. Chiropr. Med.* **2016**, *15*, 155–163. [[CrossRef](#)] [[PubMed](#)]
27. Schoenfeld, B.J.; Ratamess, N.A.; Peterson, M.D.; Contreras, B.; Tiryaki-Sonmez, G. Influence of Resistance Training Frequency on Muscular Adaptations in Well-Trained Men. *J. Strength Cond. Res.* **2015**, *29*, 1821–1829. [[CrossRef](#)]
28. Nunes, J.P.; Costa, B.D.V.; Kassiano, W.; Kunevaliki, G.; Castro-e-Souza, P.; Rodacki, A.L.F.; Fortes, L.S.; Cyrino, E.S. Different Foot Positioning During Calf Training to Induce Portion-Specific Gastrocnemius Muscle Hypertrophy. *J. Strength Cond. Res.* **2020**, *34*, 2347–2351. [[CrossRef](#)]
29. Stock, M.S.; Mota, J.A.; DeFranco, R.N.; Grue, K.A.; Jacobo, A.U.; Chung, E.; Moon, J.R.; DeFreitas, J.M.; Beck, T.W. The Time Course of Short-Term Hypertrophy in the Absence of Eccentric Muscle Damage. *Eur. J. Appl. Physiol.* **2017**, *117*, 989–1004. [[CrossRef](#)]
30. Abe, T.; Loenneke, J.P.; Thiebaud, R.S.; Loftin, M. Morphological and Functional Relationships with Ultrasound Measured Muscle Thickness of the Upper Extremity and Trunk. *Ultrasound J. Br. Med. Ultrasound Soc.* **2014**, *22*, 229–235. [[CrossRef](#)]
31. Cronin, N.J.; Finni, T.; Seynnes, O. Fully Automated Analysis of Muscle Architecture from B-Mode Ultrasound Images with Deep Learning. *arXiv* **2020**, arXiv:2009.04790.