

RESEARCH ARTICLE

Reliability of ultrasonographic measurement of muscle architecture of the gastrocnemius medialis and gastrocnemius lateralis

Samantha May^{1†}, Simon Locke¹, Michael Kingsley^{2,3*}

1 La Trobe Rural Health School, La Trobe University, Bendigo, Victoria, Australia, **2** Holsworth Research Initiative, La Trobe University, Bendigo, Victoria, Australia, **3** Department of Exercise Sciences, University of Auckland, Auckland, New Zealand

† Current address: Department of Sports Medicine, Alphington Sports Medicine Clinic, Northcote, Victoria, Australia

* M.Kingsley@latrobe.edu.au, Michael.Kingsley@Auckland.ac.nz



OPEN ACCESS

Citation: May S, Locke S, Kingsley M (2021) Reliability of ultrasonographic measurement of muscle architecture of the gastrocnemius medialis and gastrocnemius lateralis. *PLoS ONE* 16(9): e0258014. <https://doi.org/10.1371/journal.pone.0258014>

Editor: Emiliano Cè, Università degli Studi di Milano, ITALY

Received: June 16, 2021

Accepted: September 15, 2021

Published: September 29, 2021

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0258014>

Copyright: © 2021 May et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data underlying the results presented in the study are available from: <https://doi.org/10.26181/611a2abef0c7e>.

Abstract

Ultrasonography is widely used to measure gastrocnemius muscle architecture; however, it is unclear if values obtained from digitised images are sensitive enough to track architectural responses to clinical interventions. The purpose of this study was to explore the reliability and determine the minimal detectable change (MDC) of gastrocnemius medialis (GM) and gastrocnemius lateralis (GL) muscle architecture using ultrasound in a clinical setting. A trained sonographer obtained three B-mode images from each of the GM and GL muscles in 87 volunteers (44 males, 43 females; 22±9 years of age) on two separate occasions. Three independent investigators received training, then digitised the images to determine intra-rater, inter-rater, and test-retest reliability for fascicle length (FL), pennation angle (θ) and muscle thickness. Median FL, θ , and muscle thickness for GM and GL were 53.6–55.7 mm and 65.8–69.3 mm, 18.7–19.5° and 11.9–12.5°, and 12.8–13.2 mm and 15.9–16.9 mm, respectively. Intra- and inter-rater reliability of manual digitisation was excellent for all parameters. Test-retest reliability was moderate to excellent with intraclass correlation coefficient (ICC) values ≥ 0.80 for FL, ≥ 0.61 for θ , and ≥ 0.81 for muscle thickness, in both GM and GL. The respective MDC for GM and GL FL, θ , and muscle thickness was ≤ 12.1 mm and ≤ 18.00 mm, $\leq 6.4^\circ$ and $\leq 4.2^\circ$, and ≤ 3.2 mm and ≤ 3.1 mm. Although reliable, the relatively large MDC suggest that clinically derived ultrasound measurements of muscle architecture in GM and GL are more likely to be useful to detect differences between populations than to detect changes in muscle architecture following interventions.

Introduction

Skeletal muscle architecture is described by the geometrical arrangement of muscle fibre bundles relative to the force-generating axis of the muscle [1]. The arrangement influences contraction velocity and force-generating capacity of a muscle, as well as the range of reciprocating motion, known as excursion [1–5]. In pennate muscles, fibre bundles known as

Funding: The authors received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

fascicles, insert obliquely into aponeuroses. The angle derived from the insertion of a fascicle into an aponeurosis defines the pennation angle (θ), and the distance between aponeuroses defines anatomical muscle thickness. These parameters of skeletal muscle architecture have been measured in studies of muscle physiology and biomechanics to determine anatomical and contractile properties of muscles [2, 3, 6–10]. Fascicle length (FL), θ , and muscle thickness, can be measured through in vivo, two-dimensional (2D) B-mode ultrasound imaging [11–14]. B-Mode ultrasound is non-invasive, readily available in research and clinical settings, and provides superior spatial resolution of images obtained in real time. This modality provides detailed visualisations of hypoechoic fibre bundles with hyperechoic septations within the muscle, separated by hyperechoic fibroadipose tissue [3, 7, 8, 15–18]. Once a single image frame or set of images are acquired, measurements are traditionally made using manual digitisation with custom-written computer software [2, 8, 19–27] or by using automatic tracking [28–31].

Several limitations should be considered when using ultrasound to measure skeletal muscle architecture. Ultrasound is operator-dependent, and potential sources of error stem from probe placement or location, probe pressure, and probe orientation [32, 33]. Freehand probe placement allows for manual adjustments of the probe parallel to the fascicle plane to obtain the clearest possible image, although this ‘field-based’ approach is subject to operator bias and can reduce reliability of repeated measurements on the same subject [2, 32–38]. Misalignment of the probe plane with respect to the fascicle plane can lead to over- or under-estimation of FL and θ [32, 37]. Excessive probe pressure can lead to tissue deformation and reduce the accuracy of measurements [32]. Standard configurations, which are most cost-effective, include a transducer sized between 4 to 6 cm. These transducer field of view dimensions are often too narrow to simultaneously view both a fascicle’s origin on one aponeurosis and its insertion on the other. Specialised long transducers (10 cm) can be employed; however, these transducers are more costly and increase susceptibility to uneven probe pressure spread across the underlying tissues that can cause deformation [39, 40]. More advanced extended-field-of-view techniques, where entire fascicles are imaged by moving the probe with a continuous scan, increase the risk of technical error during digital reconstruction of the image [41]. In cases where the above techniques are not available and fascicles extend beyond the ultrasound’s field of view, the external segment can be estimated by linear extrapolation of the fascicle to its intersection with the aponeuroses [2, 24, 25, 42]. Linear extrapolation assumes a fascicle follows a linear path and does not account for fascicle or aponeurosis curvature, which has important implications for the accurate calculation of architecture [43, 44]. Consequently, the reproducibility of conventional B-mode ultrasound measures of skeletal muscle architecture requires investigation.

Authors of a systematic review found that, with appropriate operator training, ultrasound measures of muscle architecture are highly reliable and valid [11]. However, most of the studies in the review described strict protocols to reduce operator bias and error, used expensive machines, or used equipment that cannot be transported to other relevant sites where sports medicine services are engaged, such as athletic clubrooms [45]. The procedures that standardise: (1) positioning of the subject and ultrasound transducer, (2) scan parameters, (3) image processing, and (4) digitisation analysis are time consuming, which poses a challenge to the suitability of the use of ultrasound in a clinical setting where time and equipment are often limited [40]. Therefore, the reproducibility of findings in a controlled environment might not translate to a clinical setting, where high resolution stationary ultrasound machines or retest aids are not universally available. Exploration of the reproducibility of an ultrasound method applicable to a clinical sports medicine context has not been performed. Regardless of this,

conventional B-mode ultrasound imaging remains the most widely used modality to measure skeletal muscle architecture in clinical settings due to its practicality and affordability [39].

The gastrocnemius medialis (GM) muscle architecture has been determined to relate closely to force production capacity and ankle joint kinematics during gait [25]. The muscle is located superficially relative to skin [3, 11], and contains shorter fascicles that can be measured with a single commercial transducer sized 4–6 cm. Because of the GM's locomotor significance and ease of identification of muscle architecture under ultrasound, it has been frequently targeted in interventional studies aimed at improving biomechanical function, as well as in clinical studies of muscle disorders [46], and immobilisation studies exploring bed rest [25] or exposure to microgravity [47]. The GM combines with the gastrocnemius lateralis (GL) as well as the soleus combine to form the triceps surae muscle group, the dominant plantar flexors of the ankle [48]. The gastrocnemius muscles have different architectural properties; GL is characterised by longer FLs and smaller θ s whereas the GM comprises shorter FLs and larger θ s [7, 47].

The reliability of measuring GM architecture using ultrasonography has been reported as being between good to excellent for FL (ICC = 0.81–0.99) and θ (ICC = 0.85–1.00) following a review of studies [11] where researchers applied strict protocols with step-by-step instructions [2, 8, 24, 25, 42, 43, 46, 49]. Across these studies, different measures were repeated for the purposes of reporting reliability, and sample sizes were relatively small, which might compromise the conclusions that can be drawn. Authors analysed the reliability of image acquisition, image analysis, or both of these steps with repeated measurements, however these were performed either within a single session, across multiple sessions, or across multiple investigators [11]. These variations have made it difficult for inter-study comparisons of the values reported. When simplified to the step of image analysis alone, specifically the method of manual digitisation, studies have reported excellent reliability following a single investigator repeating the digitisation step on two occasions (intra-rater), or between multiple investigators (inter-rater) [2, 8, 11, 24, 25, 32, 42, 43, 46, 49]. However, when investigating responses to interventions, it is more valuable to determine whether the method of both image acquisition and digitisation repeated on multiple occasions is reliable and sensitive enough to track adaptations.

Historically, GM measurements are more vulnerable to error when measured between sessions [11]. Two studies revealed large discrepancies in GM FL and θ results after being scanned with ultrasound twice, and recommended caution regarding small changes in FL and θ in clinical studies as the relevant changes were likely smaller than the measurement error [32, 50]. Existing literature does not address whether the magnitude of measurement error of ultrasonography exceeds the potential changes in FL, θ , and muscle thickness following clinical interventions such as immobilisation, chronic exercise training, or disuse. The reliability of imaging the GL is far less known [11], and it cannot be assumed that the method of ultrasound imaging for GM will prove reliable for GL in the same individual or group [7, 51]. Separate analysis is required, particularly test-retest reliability estimates including the minimum detectable change. This new knowledge can inform the appropriateness of tracking architectural adaptations to both heads of the gastrocnemius muscle in individuals using ultrasound.

The primary purpose of this study was twofold: (1) to determine the reliability of ultrasound image acquisition and manual digitisation of GM and GL muscle architecture parameters in resting conditions in a clinical setting, and (2) to estimate the magnitude of difference that would be required to detect real change between or within individuals and groups. The authors hypothesised that: (a) the intra-rater and inter-rater reliability of the manual digitisation step would be excellent and align with results from previous studies, and (b) that the test-retest reproducibility of the muscle architecture results between ultrasound sessions using methods repeated in a clinical setting would be lower than previously described results in the literature

that applied stricter conditions or used equipment not readily available to the sports medicine clinician.

Methods

Study design and participants

Fifty-six adults and 31 adolescents aged between 13 and 63 (44 males and 43 females: mean \pm SD, 22 ± 9 years, 175 ± 12 cm, and 73 ± 16 kg) provided signed consent and attended two data collection sessions at least 7 days apart. An additional 2 adult and 4 adolescent participants attended a single data collection session but did not attend the second data collection session and were not included in the test-retest reliability analysis. For participants aged <18 years, written informed consent was given by a parent or legal guardian. Participants were screened for eligibility using a questionnaire. The questionnaire covered demographic information, such as age, gender, previous lower limb injuries, and activity levels. Participants were excluded if they were aged 12 or younger, had a history of an acute or chronic lower limb injury within the previous 12 months, or if they had exercised their calf muscles earlier on the day of the assessment (prolonged walking, jogging, running, sprinting, hopping, skipping, jumping, or performing heel raises). The study was approved by La Trobe University Human Ethics Committee (Reference: S17-114) and was conducted according to the National Statement on Ethical Conduct in Human Research.

Procedures

Anthropometry was undertaken in accordance with procedures described by the International Society for the Advancement of Kinanthropometry [52]. Body mass was recorded to the nearest 0.1 kg using a portable analogue floor-scale (Model 762; Seca, Germany), and stretch stature was measured to the nearest 0.1 cm using a portable stadiometer (Model 213; Seca; Germany). The muscle architecture of both left and right GM and GL muscles were assessed during resting conditions to reduce fascicle curvature [2, 43]. B-mode ultrasound images were taken from each participant on two occasions using a portable ultrasound (LOGIQ V2; GE Healthcare, Australia) with a 38 mm wide linear probe and a standardised frequency of 12–13 MHz. This portable machine was selected as this is attractive for clinical use given its lower cost and smaller size compared to large, stationary, high-precision machines [53].

Participants lay prone on an examination table with the lower leg supported on an inclined foam wedge angled at 20° so that the knee was flexed to within the desired range of 20 – 30° [54]. The wedge was consistently placed in the same position with the thin edge adjacent to the tibiofemoral joint lines to ensure the retest position would be maintained; the knee angle was confirmed on the first 20 participants with a manual goniometer. The gastrocnemius bridges both the knee and ankle joint and is under full tension when the knee is extended because the muscle origin is furthest from its insertion [54]. Knee flexion of at least 20° eliminates the effect of ankle dorsiflexion restraining the gastrocnemius, and ankle dorsiflexion and gastrocnemius tension remains unchanged between 20° and 75° of knee flexion [54]. The ankle was secured at approximately 90° using a night splint and confirmed with a manual goniometer (Fig 1). A trained sonographer used a measuring tape to find the initial probe site at one-third of the distance from the popliteal crease of the knee to the tip of the medial malleolus for the GM and the lateral malleolus for the GL, at the mid-muscle belly which was determined via inspection and palpation [7, 27, 32, 47, 49, 53, 55]. The sonographer applied ample conductive gel to ensure acoustic enhancement and reduce probe pressure against the skin. The probe was then orientated by freehand to the plane of fascicles that were visibly inserting into either the superficial or deep aponeurosis, and where the aponeuroses were close to parallel [23, 33, 37].



Fig 1. Position of lower limb during ultrasonography.

<https://doi.org/10.1371/journal.pone.0258014.g001>

This positioning around the initial placement site appreciates learned clinical practice where a structure is targeted and visualised in real time using pattern recognition [56, 57]. Depth of scanning was manually adjusted to 3–6 cm from the surface of the transducer, until both outer borders of the superficial and deep aponeuroses were visible. Multiple ultrasound images were obtained at the site where the muscle belly was widest, where FLs are thought to be homogeneously distributed [8, 58]. The probe was lifted between image captures to ensure each muscle's architecture was represented by several different measures of FL, θ and muscle thickness. This resulted in three ultrasound images captured from each GM and GL at the left and right limb on the same day within one session, totalling 12 images per participant. This process was repeated at a second session held at least 7 days later, resulting in a total of 24 images captured per participant.

Images underwent de-identification and randomisation (<https://www.randomizer.org>). The following anatomical structures were digitised frame-by-frame (Fig 2), using software designed in LabVIEW (version 16; National Instruments, USA): (1) superficial aponeurosis external and internal borders; (2) deep aponeurosis external and internal borders; and (3) proximal and distal endpoints of three muscle fascicles within the imaging plane per image. FL was defined by the length of a linear fascicle path between the superficial aponeurotic junction and the deep aponeurotic junction. Investigators received one hour of training and were provided with step-by-step training manuals that instructed investigators to prioritise selection of fascicles that were straight with visible endpoints at the junction of either the deep or superficial aponeurosis. Where the fascicles extended beyond the captured image, the software extrapolated to the aponeuroses and estimated FL, like that described in model V by Ando et al. [59]. The intersection of the line of the selected fascicle inserting obliquely with the deep aponeurosis defined θ . An average of these three measurements was used to represent FL and θ [60]. Muscle thickness was defined as the mean vertical distance between the internal borders

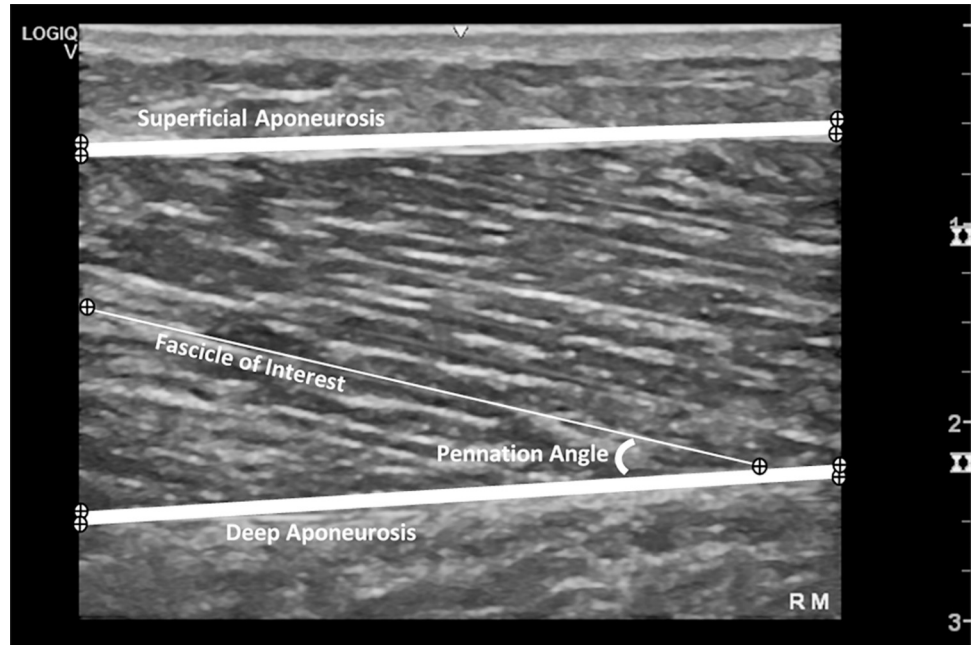


Fig 2. Ultrasound image from a right medial head of gastrocnemius of a participant, showing an example of the points plotted (⊕) for superficial aponeurosis borders, deep aponeurosis borders, and a selected fascicle length and corresponding pennation angle.

<https://doi.org/10.1371/journal.pone.0258014.g002>

of the superficial and deep aponeuroses measured on either side of the image. One investigator reviewed all 2184 de-identified images and removed 39 images (1.7%) due to poor quality, determined by inability to visualise the borders of deep aponeurosis (incorrect depth), or an unacceptable degree of curvature in the aponeuroses.

Fig 3 describes the methods used to determine reliability for the steps of the procedure. Intra-rater reliability of the digitisation method was analysed using one blinded investigator (Investigator A) repeating digitisation on 100 randomly sampled images on two occasions. Three blinded investigators (A, B and C) analysed all 2145 images to provide comparisons for inter-rater reliability of the digitisation method. The images were divided into images of GM (1081 images) or GL (1064 images), and reliability analyses were performed for these groups separately. For test-retest reliability of the ultrasound image acquisition and digitisation method combined, results were re-identified for each of the 87 participants and separated into

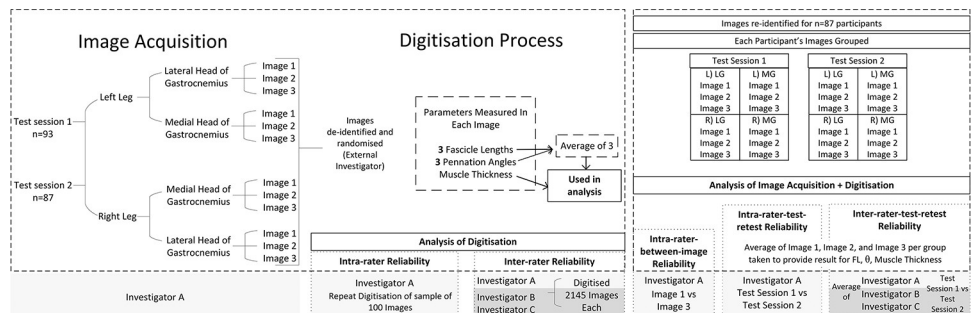


Fig 3. Protocols for reliability analyses.

<https://doi.org/10.1371/journal.pone.0258014.g003>

image number (1, 2 or 3) and test session (1 or 2). Intra-rater between-image reliability was measured by comparing Investigator A's results from session 1 image 1, to session 1 image 3. Intra-rater-test-retest reliability was determined by comparing Investigator A's averages from test session 1 images to test session 2. The average of the three investigators (A, B and C) results from test session 1 was also compared against their average of results from test session 2 to determine inter-rater-test-retest reliability.

Statistical analysis

FL, θ , and muscle thickness were assessed for normality using the Shapiro-Wilk test, alongside assessment of skewness, kurtosis, and distribution curves. The majority of data breached the assumption of normality ($p < 0.05$), and group data are presented as median and inter-quartile range (IQR).

Relative consistency was quantified using intraclass correlation coefficients (ICC) according to guidelines defined by McGraw and Wong (1996) with 95% confidence intervals (95% CI) [61]. The ICC values were calculated for intra-rater (ICC; two-way mixed-effects model, absolute definition, single rater type), inter-rater (ICC; two-way random-effects model, absolute definition, single rater type), and test-retest reproducibility (two-way mixed-effects model, absolute definition, average rater type) [61, 62]. ICC values were classified according to Koo and Li (2016) as excellent (≥ 0.90), good (0.75 to 0.89) moderate (0.50 to 0.74), and poor (≤ 0.49) [62]. Absolute reliability was determined using the methods of Bland and Altman (1999) [63]. Heteroscedasticity was examined through visual inspection of the Bland-Altman plots and the correlation between the absolute differences and their respective means. The correlation was significant for the majority of comparisons, and data were log transformed prior to further analysis of the agreement between measures [63]. Consequently, mean bias and 95% ratio limits of agreement (RLOA) relative to the mean were calculated to provide an indication of systematic error and random error, respectively. These values show how much two scores varied between investigators or between sessions; mean bias values closer to 0 indicate higher reliability, and narrower RLOA are preferable [64].

Test-retest reliability was further assessed by calculating the standard error of the measurement (SEM), as follows: $SEM = SD\sqrt{1 - ICC}$, where SD is the standard deviation of the measure [65]. The minimal detectable change (MDC) defines the smallest change in the scores between two sessions that needs to be observed in order to be 95% confident that the observed change in muscle architecture is real and not a product of measurement error or random variation in the method [66]. The SEM was used to calculate the MDC with the equation:

$MDC = SEM \times \sqrt{2} \times z_{score}_{(95\% CI)}$, where the z_{score} equals 1.96, and the multiplier of $\sqrt{2}$ is to account for the uncertainty introduced by using difference scores from measurements at 2 points in time [67, 68]. In addition, Cohen's d effect size, measures the proportion of variance shared by groups, and was calculated as the difference in group mean values divided by the square root of the pooled variance (population SD) [69]. An effect size of 0.20 was considered to be small, 0.50 medium, and 0.80 large [70]. Results were considered significant at $p \leq 0.05$, and statistical analyses were performed using IBM SPSS (V25.0; Armonk, NY: IBM Corp.) [71].

Results

Analysis of manual digitisation

Intra-rater reliability. Intra-rater reliability was excellent for FL (ICC = 0.99, 95% CI: 0.98–0.99), θ (ICC = 0.99, 95% CI: 0.98–0.99), and for muscle thickness (ICC = 0.99, 95% CI:

Table 1. Gastrocnemius medialis and gastrocnemius lateralis median (IQR) values and inter-rater reliability results for fascicle length, pennation angle and muscle thickness lengths.

Gastrocnemius Medialis (1081 images)				Gastrocnemius Lateralis (1064 images)			
Median (IQR)	ICC (95% CI)	Mean Bias (RLOA)	Cohens	Median (IQR)	ICC (95% CI)	Mean Bias (RLOA)	Cohens
<i>Fascicle Length</i>							
54.6 mm (49.1–62.9)	0.95 (0.95 to 0.96)	-0.5% (-12.3 to 12.8)	0.07	67.6 mm (59.6 to 77.3)	0.91 (0.90 to 0.92)	-0.8% (-16.5 to 17.7)	0.08
<i>Pennation Angle</i>							
19.1° (16.4–21.5)	0.95 (0.94 to 0.96)	0.3% (-11.4 to 13.5)	0.05	12.2° (10.1 to 13.9)	0.94 (0.94 to 0.95)	1.5% (-13.6 to 19.2)	0.18
<i>Thickness</i>							
16.4 mm (14.6–18.7)	1.00 (1.00 to 1.00)	0.1% (-3.1 to 3.4)	0.07	13.0 mm (11.0 to 14.8)	0.99 (0.99 to 1.00)	0.2% (-3.9 to 4.5)	0.11

IQR = interquartile range; ICC = intra-class correlation coefficient; CI = confidence interval; RLOA = ratio limits of agreement, 95% CI; mm = millimetres; ° = degrees.

<https://doi.org/10.1371/journal.pone.0258014.t001>

0.99–0.99). Systematic bias was small, with mean differences (MD) of 0.3%, 0.7% and 0.1%, respectively. In absolute values, MD were 0.3 mm, 0.1°, and 0.1 mm. The RLOA were -8.1% to 8.0% for FL, -9.4% to 11.9% for θ , and -2.7% to 3.0% for muscle thickness.

Inter-rater reliability. Table 1 displays median (IQR) values and inter-rater reliability analyses for the acquired images of GM and GL. For all comparisons, the ICC values were excellent (≥ 0.91), with the widest 95% CI being 0.90–0.92. The ICC values were strongest for muscle thickness comparisons, with ICC values of 0.99 to 1.00. Mean differences between Investigators A, B and C were small in magnitude for both GM and GL. For all parameters, there was no evidence of systematic bias (Fig 4), and the mean bias ranged from 0.1% to 0.5% for GM and 0.2% to 1.5% for GL. For example, the mean bias for GM FL equates to small absolute values (≤ 0.3 mm) in proportion to the median FL of 54.6 mm. For GL FL, the difference in absolute values was again small (≤ 0.5 mm) against a median FL of 67.6 mm. The GL θ showed the largest systematic difference which was found between Investigators A and B: MD: 2.2% (RLOA: -12.1%– 18.7%), as well as between Investigators A and C being MD: 2.2% (RLOA: -13.3%– 20.5%) (S1 Table). This equates to 0.1° for both, corresponding to 0.8% of the median angle. For both muscles, muscle thickness was the most reliable parameter measured with the tightest RLOA (-3.1% to 3.4% for GM and -3.9% to 4.5% for GL). Cohens d statistic was small (< 0.20) in all comparisons.

Analysis of image acquisition and manual digitisation

Intra-rater between image reliability. Relative reliability was excellent for FL (ICC = 0.95, 95% CI: 0.94–0.96), θ (ICC = 0.94, 95% CI: 0.93–0.95), and for muscle thickness (ICC = 0.97, 95% CI: 0.97–0.98). Systematic bias between image 1 to image 3 was small, with mean differences (MD) of 0.2%, 1.1% and 0.1% respectively. The RLOA were wide, within the range of -14.9% to 17.2% for FL, -19.1% to 26.2% for θ , and 10.0% to 10.8% for muscle thickness.

Test-retest reliability. The results of the image acquisition and digitisation inter-rater-test-retest and intra-rater-test-retest analyses are presented in Table 2. GM inter-rater and intra-rater comparisons for FL was the most reliable measure, with good to excellent relative reliability (ICC range: 0.88 to 0.92). ICCs were moderate to good for θ s with the widest values ranging from 0.63 to 0.76, and good for muscle thickness (ICC range: 0.83 to 0.86). Similar findings were seen for GL (Table 2); FL was less reliable (ICC range: 0.80 to 0.84). Mean results and MDC values were similar between intra-rater and inter-rater analyses. Mean bias was $\leq 3.6\%$ for all parameter comparisons in both muscles, and there were no systematic differences between the left limb and right limb. The left limb results are represented in the

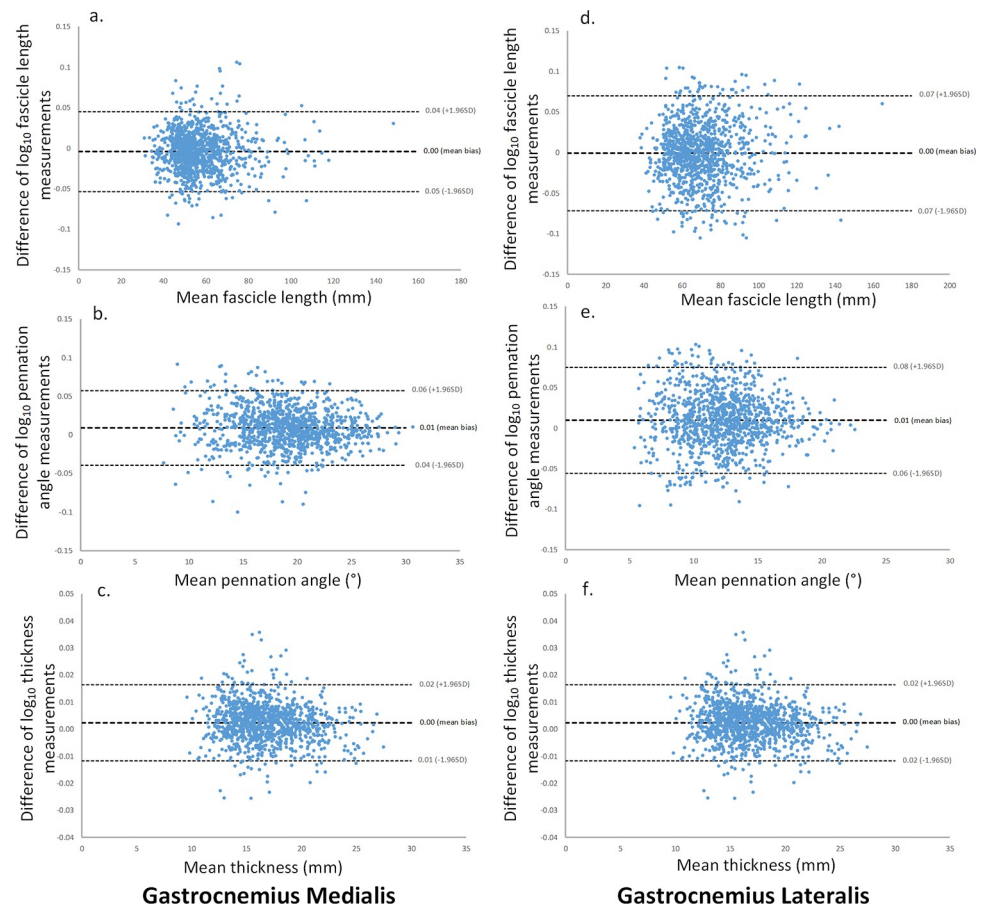


Fig 4. Bland-Altman plots showing the differences between Investigator A and Investigator B \log_{10} measurements in relation to the mean of Investigator A and Investigator B measurements. Dashed lines indicate the mean bias and the upper and lower 95% limits of agreement. Fascicle length, pennation angle, and muscle thickness measurement comparisons for the gastrocnemius medialis are represented by a, b, and c, respectively, and the gastrocnemius lateralis represented by d, e, and f.

<https://doi.org/10.1371/journal.pone.0258014.g004>

Bland-Altman plots in Fig 5. The mean bias was close to zero for all parameters. In contrast, the 95% RLOA were wide, demonstrating random error in the method. In terms of precision, absolute reliability was stronger when measuring muscle thickness in both the GM (SEM = 1.1–1.2; MDC = 3.2 mm) and the GL muscles (SEM = 0.9–1.6; MDC = 2.6–3.1 mm). The largest inter-rater GM FL MDC was 12.1 mm, which represents 22.0% the median 54.9 mm. For the GL, the greatest inter-rater FL MDC was 17.3 mm, or 25.9% of the associated FL median of 66.7 mm. Cohens values were < 0.29 for all parameter comparisons.

Discussion

The findings confirm that reliability is excellent when using manual digitisation to measure gastrocnemius FL, θ and muscle thickness on images derived from ultrasound, within a single investigator on two occasions, and between multiple investigators (ICCs \geq 0.91). Inter-rater digitisation results for the GL were slightly less reliable when compared to the GM, likely due to the anatomical differences between the two muscles. Reliability is reduced when ultrasound image acquisition is repeated within a session which required the probe to be removed and replaced between scans. Reliability is further reduced when participants attend re-scanning on

Table 2. Test-retest reliability for fascicle length, pennation angle and muscle thickness of the gastrocnemius medialis and gastrocnemius lateralis muscles.

Gastrocnemius Medialis (n = 87)							
	Test1 Median (IQR)	Test2 Median (IQR)	ICC (95% CI)	Mean Bias (RLOA)	SEM	MDC	Cohens
Inter-rater-Test-retest							
<i>Fascicle Length</i>							
Left	55.7 mm (49.5–62.3)	54.0 mm (49.5–61.5)	0.88 (0.81–0.92)	1.5% (-21.7–31.6)	4.4	12.1	0.11
Right	53.7 mm (49.3–63.6)	54.6 mm (49.6–62.6)	0.92 (0.87–0.94)	0.1% (-18.9–23.5)	3.5	9.8	0.03
<i>Pennation Angle</i>							
Left	18.7° (15.8–21.5)	19.4° (16.8–22.2)	0.75 (0.62–0.84)	-3.4% (-34.4–42.1)	2.1	5.9	0.15
Right	19.0° (15.8–20.7)	19.1° (16.8–21.1)	0.64 (0.45–0.77)	-3.4% (-36.2–46.2)	2.3	6.3	0.14
<i>Thickness</i>							
Left	15.9 mm (14.0–18.1)	16.5 mm (14.9–18.7)	0.86 (0.78–0.91)	-2.7% (-24.0–24.6)	1.1	3.2	0.22
Right	16.3 mm (14.5–18.3)	16.8 mm (14.9–18.6)	0.83 (0.74–0.89)	-3.4% (-25.6–25.5)	1.2	3.2	0.27
Intra-rater-Test-retest							
<i>Fascicle Length</i>							
Left	55.4 mm (48.8–62.3)	54.2 mm (48.8–62.0)	0.88 (0.82–0.92)	1.2% (-21.8–30.9)	4.3	12.0	0.09
Right	53.6 mm (48.5–62.7)	54.1 mm (48.8–61.9)	0.91 (0.86–0.94)	-0.1% (-20.1–24.9)	3.7	10.3	0.01
<i>Pennation Angle</i>							
Left	18.7° (16.3–21.9)	19.5° (17.0–22.3)	0.76 (0.63–0.84)	-2.9% (-33.2–41.3)	2.1	5.7	0.13
Right	19.2° (16.3–20.9)	19.1° (17.2–21.3)	0.63 (0.43–0.76)	-3.2% (-36.1–46.6)	2.3	6.4	0.13
<i>Thickness</i>							
Left	15.9 mm (14.1–18.2)	16.5 mm (14.9–18.9)	0.86 (0.78–0.91)	-2.8% (-24.1–24.6)	1.1	3.2	0.23
Right	16.4 mm (14.5–18.3)	16.9 mm (15.0–18.7)	0.83 (0.73–0.89)	-3.6% (-25.8–25.2)	1.2	3.2	0.29
Gastrocnemius Lateralis (n = 87)							
	Test1 Median (IQR)	Test2 Median (IQR)	ICC (95% CI)	Mean Bias (RLOA)	SEM	MDC	Cohens
Inter-rater-Test-retest							
<i>Fascicle Length</i>							
Left	68.4 mm (59.5–77.2)	68.7 mm (62.4–77.3)	0.84 (0.75–0.89)	-2.0% (-26.3–30.5)	5.7	15.9	0.13
Right	67.1 mm (59.5–77.8)	66.3 mm (58.0–73.4)	0.80 (0.70–0.87)	2.8% (-25.6–42.0)	6.2	17.3	0.15
<i>Pennation Angle</i>							
Left	12.3° (9.8–13.6)	11.9° (9.6–13.7)	0.79 (0.68–0.87)	3.6% (-21.3–56.2)	1.4	3.8	0.16
Right	12.0° (10.9–14.0)	12.3° (10.5–13.9)	0.61 (0.41–0.75)	-1.7% (-37.1–53.8)	1.5	4.2	0.09
<i>Thickness</i>							
Left	12.8 mm (11.0–14.3)	12.8 mm (10.4–14.5)	0.88 (0.82–0.92)	1.3% (-22.9–33.2)	0.9	2.6	0.06
Right	13.2 mm (11.4–15.0)	12.8 mm (10.9–14.5)	0.81 (0.71–0.88)	3.2% (-23.0–38.5)	1.6	3.1	0.20
Intra-rater-Test-retest							
<i>Fascicle Length</i>							
Left	67.3 mm (59.8–76.9)	69.3 mm (61.4–76.5)	0.82 (0.73–0.89)	-1.9% (-27.4–32.5)	5.9	16.5	0.13
Right	67.1 mm (59.2–78.8)	65.8 mm (57.2–74.3)	0.80 (0.69–0.87)	2.0% (-27.4–43.3)	6.5	18.0	0.08
<i>Pennation Angle</i>							
Left	12.4° (10.0–13.8)	12.1° (9.8–14.1)	0.79 (0.67–0.86)	3.3% (-32.3–57.4)	1.4	3.9	0.14
Right	12.4° (10.9–14.2)	12.5° (10.8–14.1)	0.63 (0.43–0.76)	-0.9% (-36.4–54.4)	1.5	4.1	0.06
<i>Thickness</i>							
Left	12.8 mm (11.0–14.3)	12.8 mm (10.5–14.8)	0.88 (0.81–0.92)	1.2% (-23.1–33.2)	0.9	2.6	0.05
Right	13.2 mm (11.6–15.0)	12.9 mm (11.0–14.5)	0.81 (0.71–0.88)	3.3% (-22.9–38.3)	1.6	3.1	0.21

n = participant number; Test1 = first test session; IQR = interquartile range; Test2 = second test session; ICC = intra-class correlation coefficient; CI = confidence interval; RLOA = 95% ratio limits of agreement; SEM = standard error of measurement; MDC = minimal detectable change; mm = millimetres; ° = degrees.

<https://doi.org/10.1371/journal.pone.0258014.t002>

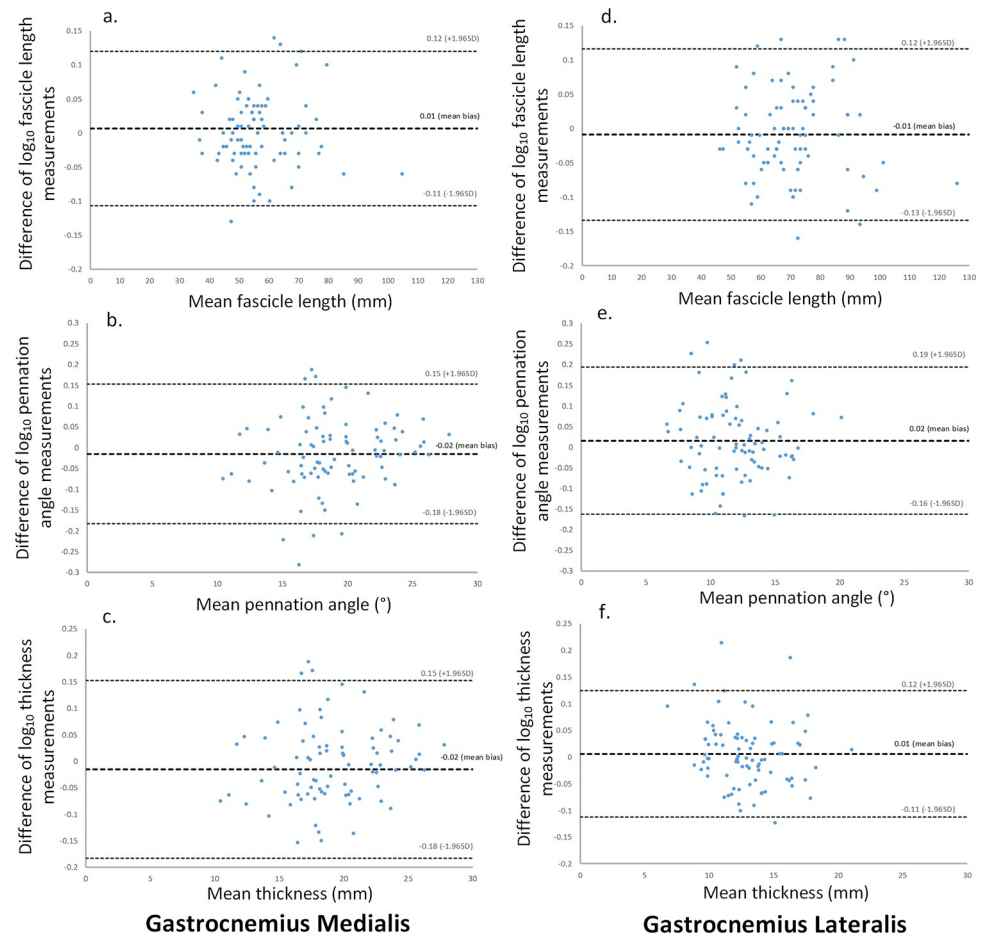


Fig 5. Inter-rater-test-retest Bland-Altman plots of the differences of log-transformed measurements between test session 1 and test session 2 in relation to the mean of test session 1 and test session 2 measurements. Dashed lines indicate the mean bias and the upper and lower 95% limits of agreement. Fascicle length, pennation angle, and muscle thickness measurement comparisons for the gastrocnemius medialis are represented by a, b, and c, respectively, and the gastrocnemius lateralis represented by d, e, and f.

<https://doi.org/10.1371/journal.pone.0258014.g005>

a second occasion. Therefore, this method for test-retest analysis is unlikely to accurately track changes to an individual's gastrocnemius muscle architecture in interventional studies.

Manual digitisation is a simple process that is repeatable with excellent reliability within a single investigator and between investigators, which suggests that subjectively choosing different fascicles within a single image does not reduce reliability. Selecting three fascicles and using the average of these further improves reliability. Where direct comparisons could be made with previous studies that reported inter-rater reliability of manual digitisation of the GM but only selected one single fascicle, this study's ICCs of 0.95, 0.95, and 1.00 surpassed the reported ICCs of 0.93, 0.82, and 0.96 for FL, θ , and muscle thickness, respectively [32]. Selection of three fascicles is time efficient, only requiring an additional 5–10 seconds to complete. Future studies should consider this method rather than choosing one single fascicle. Furthermore, the lack of previous digitisation experience of investigators in this study demonstrates that manual digitisation is easy to learn with a set of instructions explaining a systematic approach to identifying anatomical landmarks. The digitiser does not need prior medical training, but basic computer skills are necessary. Of all the parameters, relative and absolute reliability was strongest for muscle thickness comparisons with narrowest RLOA. This

precision could be derived from the direct pinpointing of aponeurotic borders to find measurements of thickness, rather than selecting three random fascicles out of many to calculate FL and θ readings.

Specific to the GM, findings for FL and θ are similar or stronger than previous findings measuring reliability within investigators, where ICCs were between 0.87–0.99 for FL and between 0.85–0.99 for θ [8, 24, 25, 43, 49, 53, 72]. Between multiple investigators, this study's results for FL agree with previously reported ICCs > 0.91 [11, 32, 50]. The results for θ are stronger than previously reported values between 0.72 and 0.96 [50]. It should be noted that the reliability data presented here are specific to the protocol that this study designed. This is the first study that measures reliability of manual digitisation of FL, θ and muscle thickness between multiple investigators using the averages of 3 FLs and 3 θ s from one image. The GL FLs were less reliable than the GM following repeated manual digitisation, although still excellent (ICCs > 0.95 for GM and > 0.91 for GL). This finding is likely to reflect the architectural structure of the GL muscle with longer FLs [7, 47], meaning that the ultrasound transducer field of view was often too narrow to view a long GL fascicle's origin on one aponeurosis and its insertion on the other. The transducer configuration of 4 cm to 6 cm is considered standard size and is often the cheapest and most accessible option, however when fascicles extend outside of the ultrasound image, estimating the length of the missing portion of fascicle can increase the random error in both FL and θ calculations. In specific muscles where fascicle lengths are longer, such as the GL or those in the thigh, this limitation can be overcome with specialised long transducers (10 cm) or extended-field-of-view techniques, however increased cost or time required to learn and reliably perform this technique poses some additional challenges to the clinician.

Specific to absolute test-retest reliability analyses of combined image acquisition and manual digitisation, this study found wide RLOA, indicating a large degree of variability in the method particularly when a participant returned on a separate occasion for a repeat scan. While there is some variability due to manual manipulation of the probe as it is replaced and repositioned on a participant within a single scanning session, random error was most significant when a participant was 'reset' between sessions. Skin markers, acetate paper, or plastic sheets have been used as retest aids in previous studies [25, 27, 73]. Nevertheless, this study preferentially used a system of using landmarks to set probe location, with the sonographer visually adapting the probe to obtain the best quality images per participant. Additional factors such as the amount of recent exercise to the calf, the participant inadvertently changing knee angle by self-shifting leg position on the examination table against the carefully placed foam wedge, or the size of foot in the splint impacting the angle of the ankle joint are all potential contributors to the variability between sessions and the resulting test-retest reliability. However, the small differences in the mean bias for all parameters, as well as small effect sizes between the compared groups indicates that there is no systematic error in the method used for image acquisition or manual digitisation. In terms of relative reliability, the test-retest findings align with past studies that measured GM FL and muscle thickness using protocols that use retest aids (ICC > 0.81) [8, 42, 46, 49, 53]. The comparative strength in relative reliability without the use of retest aids could be explained by the uniqueness in the current method in that the average of three images was used for session one and then compared to the average results from three images from session two. Measurements of θ were less consistent, with wider RLOA between sessions and test-retest reliability scores between moderate to good which is lower than previously reported for the GM [8, 42, 46, 49, 53]. As the analysis of manual digitisation within and between investigators revealed strong results across the three parameters, and the intra-rater between-image reliability for θ remained excellent, it can be pinpointed that θ measurements are more greatly affected by error from re-positioning of

participants and loss of precision of the location of the probe between sessions, rather than transducer orientation and rotation.

This study found that the magnitude of error of the measurement between sessions outweighs the small changes to muscle architecture that are typically seen when clinically relevant interventions are applied to gastrocnemius muscles in individuals. The changes that have been previously described to the GM muscles following interventions are frequently reported in percentages in past studies. Comparisons to previous studies can be drawn, because the descriptive measurements (median + IQR) obtained agree well with previous measurements of GM architecture in healthy adults using B-mode ultrasonography [2, 25, 35, 42, 72, 74–76]. Historically, changes seen within GM muscles following interventions are typically less than 20% for muscle thickness and even smaller values for fascicle length. One such study performed in a clinical setting found that changes to muscle thickness following 10 days of immobilisation were -15.69% in the GM, equating to approximately -2.8 mm of the mean thickness 17.9 mm (\pm 2.8 mm), although this was not statistically significant [74]. This study found that the MDC, or smallest change that can be detected beyond measurement error, was 3.2 mm for intra-rater-test-retest and inter-rater-test-retest analysis of GM muscle thickness. This corresponds to approximately 20% of the median muscle thickness width determined in this study of 16.4 mm (IQR: 14.6 mm to 18.7 mm). A further example can be made by exploring changes to the GM θ , which reportedly should be at least 6–7% [32] to be valid differences. One interventional study that concluded that after 5 weeks of bed rest, the GM θ declined by 14.3% [25]. An additional study found that disuse atrophy following injury (compared with uninjured leg; time period of injury not reported) resulted in a 12.7% decrease in FL, and a 16.4% decrease in θ [77]. A separate group of investigators measured GM fascicle length plasticity following 7 weeks of eccentric training in ten adult males, and showed that FL and θ significantly increased by average 7.6% and 6.8% respectively [21]. The MDC values determined in this study for GM θ s ranged from 5.7° to 6.4° following the intra-rater-test-retest and inter-rater-test-retest analysis, equating to a range of 29.2% to 33.3% of the median angle values. MDC values for GM FLs ranged from 9.8 mm to 12.1 mm, or to 17.9% to 22.4% of the corresponding median FLs. These MDC values indicate that the magnitude of measurement error in using the method in this study most likely exceeds the potential changes in muscle architecture in an individual following clinical intervention such as immobilisation or eccentric training. Changes within these MDC values could not be interpreted as real although changes beyond would be considered valid. Clinicians should be aware of these MDC values when interpreting results obtained using similar techniques and consider implementing more rigorous test-retest procedures if the intention is to monitor test-retest changes in muscle architecture. An alternative and appropriate use of this method in clinical practice is to compare results between populations. The absence of systematic differences represented by the narrow mean bias supports the clinical value of evaluating differences in GM architecture between adolescent and adult group populations in a clinical setting [27].

Less data is available on plasticity of GL muscle architecture following intervention, and it cannot be assumed that adaptations observed for the GM represents the GL as different mechanical loads have been observed for the different components of the triceps surae muscle [72, 78]. The descriptive analysis highlights the differences between GM and GL architecture, consolidating the findings of previous studies that have shown that there exists a relationship between pennation angle and muscle thickness [15, 79, 80]. Muscles with greater thickness have larger pennation angles, such is seen in the GM compared to the narrower GL muscle, with corresponding smaller θ s [79]. This study also confirmed that GL FLs are longer [7, 47]. One of few studies investigating changes specific to GL following training interventions found that 14 weeks of eccentric training resulted in no significant increase in FL with the change

reported to be 10%, from 86 mm \pm 18 mm to 95 mm \pm 28 mm [81]. A more recent study contrastingly found that adaptations were significant following 12 weeks of eccentric training, with FLs increasing by 8.8% [72]. Reductions in FL following lower limb immobilisation have also been observed in the lateral gastrocnemius with approximately 9% decrements in fascicle length after 23 days of lower limb suspension [60]. The test-retest analysis revealed that the MDC for GL FL ranged from 15.9 mm to 18.0 mm following the intra- and inter-rater analysis. This corresponds to a range of 23.1% to 27.3% of the average length of a fascicle. These MDC values for the GL are greater than the size of adaptations that have been reported in the literature. Therefore, the methods used in this study cannot confidently be used to observe true changes in muscle architecture in individuals following clinical intervention unless the changes were significantly larger than what has previously been reported in interventional studies. Including procedures employed between sessions to further standardise the probe location, such as including individualised limb moulds marked with anatomical landmarks, are likely to improve reproducibility but might be less pragmatic for use in clinical practice.

Additional limitations of this study should be acknowledged. A single trained sonographer obtained all the images used in the analysis, so this study was unable to assess inter-rater-test-retest reliability with multiple investigators performing both the image acquisition step as well as the manual digitisation step. The choice of 2D B-mode ultrasound to obtain data at-rest has been validated against cadaveric measurements; however, this modality does have a restricted field of view and uses mathematical equations to extrapolate lengths, as well as cannot appreciate the often curved, three-dimensional nature of a fascicle or aponeurosis. We followed the approach by previous authors to use the distance between two points of attachment to aponeuroses and to prioritise fascicles that were straight and parallel to the image plane to reduce the influence of omitting fascicle curvature on findings [82]. A simple model was designed to standardise the process and minimise error, which meant only GM and GL muscles were measured, with participants at rest in one position. The application of the reliability results to other positions, dynamic conditions, or other muscle groups is not recommended.

Conclusion

In summary, the protocol employed in this study using 2D B-mode ultrasound in a clinical setting can be used to reliably measure muscle architecture of the GM and GL muscles, in a healthy adolescent or adult population. This study presented a range of reliability statistics and divided the protocol into stages for reliability analysis that allows wider comparison to previous and future literature. Within a single investigator, repeating manual digitisation alone is most reliable, followed by obtaining measures with image acquisition combined with manual digitisation within one session, then lastly obtaining measures with image acquisition combined with manual digitisation between two sessions. Comparisons between multiple investigators undertaking manual digitisation remains reliable both between images and between sessions. However, between session measurements have a high degree of variability both within an investigator or between investigators. Therefore, the ability of this protocol to determine true change to GM or GL architecture following a clinical intervention is unreliable on an individual level, as the large MDCs and RLOAs observed are greater in magnitude than the typical changes previously reported in gastrocnemius muscles after intervention. The reliability of measurements between sessions can be improved by using stricter protocols that standardise the probe location using retest aids. Further research is required to determine whether 2D B-mode ultrasonography is sensitive enough to track adaptations to muscle architecture in individual gastrocnemius muscles in a clinical setting. However, the current protocol is useful for comparisons of two population groups and can be used for this purpose in future research.

Supporting information

S1 Table. Inter-rater reliability: Pairwise comparisons between Investigators A, B and C for fascicle length, pennation angle and muscle thickness of the gastrocnemius medialis and gastrocnemius lateralis muscles. IQR = interquartile range; ICC = intra-class correlation coefficient; CI = confidence interval; RLOA = ratio limits of agreement; mm = millimetres; ° = degrees.
(DOCX)

Acknowledgments

The authors would like to acknowledge the contribution of Bonnie Pearson, Jack Lawrence, Nikita O'Brien, Patrick Blood, Jason Canterford and Emily Densley-Walker, who assisted with the collection and analysis of ultrasound images.

Author Contributions

Conceptualization: Samantha May, Simon Locke, Michael Kingsley.

Data curation: Samantha May, Michael Kingsley.

Formal analysis: Samantha May, Michael Kingsley.

Investigation: Samantha May.

Methodology: Samantha May, Simon Locke, Michael Kingsley.

Resources: Michael Kingsley.

Software: Michael Kingsley.

Supervision: Simon Locke, Michael Kingsley.

Validation: Samantha May.

Writing – original draft: Samantha May.

Writing – review & editing: Samantha May, Simon Locke, Michael Kingsley.

References

1. Lieber RL, Friden J. Functional and clinical significance of skeletal muscle architecture. *Muscle Nerve*. 2000; 23(11):1647–66. [https://doi.org/10.1002/1097-4598\(200011\)23:11<1647::aid-mus1>3.0.co;2-m](https://doi.org/10.1002/1097-4598(200011)23:11<1647::aid-mus1>3.0.co;2-m) PMID: 11054744.
2. Narici MV, Binzoni T, Hiltbrand E, Fasel J, Terrier F, Cerretelli P. In vivo human gastrocnemius architecture with changing joint angle at rest and during graded isometric contraction. *J Physiol*. 1996; 496 (Pt 1):287–97. Epub 1996/10/01. <https://doi.org/10.1113/jphysiol.1996.sp021685> PMID: 8910216; PubMed Central PMCID: PMC1160844.
3. Fukunaga T, Ichinose Y, Ito M, Kawakami Y, Fukashiro S. Determination of fascicle length and pennation in a contracting human muscle in vivo. *J Appl Physiol* (1985). 1997; 82(1):354–8. <https://doi.org/10.1152/jappl.1997.82.1.354> PMID: 9029238.
4. Lieber RL, Friden J. Clinical significance of skeletal muscle architecture. *Clin Orthop Relat Res*. 2001; (383):140–51. Epub 2001/02/24. <https://doi.org/10.1097/00003086-200102000-00016> PMID: 11210948.
5. Hodges PW, Pengel LH, Herbert RD, Gandevia SC. Measurement of muscle contraction with ultrasound imaging. *Muscle Nerve*. 2003; 27(6):682–92. Epub 2003/05/27. <https://doi.org/10.1002/mus.10375> PMID: 12766979.
6. Wickiewicz TL, Roy RR, Powell PL, Edgerton VR. Muscle architecture of the human lower limb. *Clin Orthop Relat Res*. 1983;(179):275–83. Epub 1983/10/01. PMID: 6617027.

7. Kawakami Y, Ichinose Y, Fukunaga T. Architectural and functional features of human triceps surae muscles during contraction. *J Appl Physiol* (1985). 1998; 85(2):398–404. <https://doi.org/10.1152/jappl.1998.85.2.398> PMID: 9688711.
8. Maganaris CN, Baltzopoulos V, Sargeant AJ. In vivo measurements of the triceps surae complex architecture in man: implications for muscle function. *J Physiol*. 1998; 512 (Pt 2):603–14. Epub 1998/10/09. <https://doi.org/10.1111/j.1469-7793.1998.603be.x> PMID: 9763648; PubMed Central PMCID: PMC2231202.
9. Blazeovich AJ. Effects of physical training and detraining, immobilisation, growth and aging on human fascicle geometry. *Sports Med*. 2006; 36(12):1003–17. Epub 2006/11/25. <https://doi.org/10.2165/00007256-200636120-00002> PMID: 17123325.
10. Blazeovich AJ, Cannavan D, Coleman DR, Horne S. Influence of concentric and eccentric resistance training on architectural adaptation in human quadriceps muscles. *J Appl Physiol* (1985). 2007; 103 (5):1565–75. Epub 2007/08/25. <https://doi.org/10.1152/japplphysiol.00578.2007> PMID: 17717119.
11. Kwah LK, Pinto RZ, Diong J, Herbert RD. Reliability and validity of ultrasound measurements of muscle fascicle length and pennation in humans: a systematic review. *J Appl Physiol* (1985). 2013; 114(6):761–9. <https://doi.org/10.1152/japplphysiol.01430.2011> PMID: 23305989.
12. Scholten RR, Pillen S, Verrips A, Zwarts MJ. Quantitative ultrasonography of skeletal muscles in children: normal values. *Muscle Nerve*. 2003; 27(6):693–8. Epub 2003/05/27. <https://doi.org/10.1002/mus.10384> PMID: 12766980.
13. Oliveira VBd Carneiro SP, Oliveira LFd. Reliability of biceps femoris and semitendinosus muscle architecture measurements obtained with ultrasonography. *Research on Biomedical Engineering*. 2016; 32:365–71.
14. Timmins RG, Shield AJ, Williams MD, Lorenzen C, Opar DA. Biceps femoris long head architecture: a reliability and retrospective injury study. *Medicine and science in sports and exercise*. 2015; 47(5):905–13. <https://doi.org/10.1249/MSS.0000000000000507> PMID: 25207929.
15. Rutherford OM, Jones DA. Measurement of fibre pennation using ultrasound in the human quadriceps in vivo. *European journal of applied physiology and occupational physiology*. 1992; 65(5):433–7. Epub 1992/01/01. <https://doi.org/10.1007/BF00243510> PMID: 1425649.
16. Herbert RD, Gandevia SC. Changes in pennation with joint angle and muscle torque: in vivo measurements in human brachialis muscle. *J Physiol*. 1995; 484 (Pt 2):523–32. Epub 1995/04/15. <https://doi.org/10.1113/jphysiol.1995.sp020683> PMID: 7602542; PubMed Central PMCID: PMC1157912.
17. Narici M. Human skeletal muscle architecture studied in vivo by non-invasive imaging techniques: functional significance and applications. *J Electromyogr Kinesiol*. 1999; 9(2):97–103. Epub 1999/03/31. [https://doi.org/10.1016/s1050-6411\(98\)00041-8](https://doi.org/10.1016/s1050-6411(98)00041-8) PMID: 10098710.
18. Sartoris R, Silvestri E. *Ultrasound Basic Anatomy. Ultrasound Anatomy of Lower Limb Muscles: A Practical Guide*. Cham: Springer International Publishing; 2015. p. 23–6.
19. Aagaard P, Andersen JL, Dyhre-Poulsen P, Leffers AM, Wagner A, Magnusson SP, et al. A mechanism for increased contractile strength of human pennate muscle in response to strength training: changes in muscle architecture. *J Physiol*. 2001; 534(Pt. 2):613–23. Epub 2001/07/17. <https://doi.org/10.1111/j.1469-7793.2001.t01-1-00613.x> PMID: 11454977; PubMed Central PMCID: PMC2278719.
20. Chleboun GS, France AR, Crill MT, Braddock HK, Howell JN. In vivo measurement of fascicle length and pennation angle of the human biceps femoris muscle. *Cells Tissues Organs*. 2001; 169(4):401–9. Epub 2001/08/08. <https://doi.org/10.1159/000047908> PMID: 11490120.
21. Chleboun GS, Busic AB, Graham KK, Stuckey HA. Fascicle length change of the human tibialis anterior and vastus lateralis during walking. *J Orthop Sports Phys Ther*. 2007; 37(7):372–9. Epub 2007/08/23. <https://doi.org/10.2519/jospt.2007.2440> PMID: 17710906.
22. Gillett JG, Barrett RS, Lichtwark GA. Reliability and accuracy of an automated tracking algorithm to measure controlled passive and active muscle fascicle length changes from ultrasound. *Comput Methods Biomech Biomed Engin*. 2013; 16(6):678–87. Epub 2012/01/13. <https://doi.org/10.1080/10255842.2011.633516> PMID: 22235878.
23. Lidstone DE, van Werkhoven H, Stewart JA, Gurchiek R, Burris M, Rice P, et al. Medial gastrocnemius muscle-tendon interaction and architecture change during exhaustive hopping exercise. *J Electromyogr Kinesiol*. 2016; 30:89–97. Epub 2016/07/01. <https://doi.org/10.1016/j.jelekin.2016.06.006> PMID: 27362587.
24. Aggeloussis N, Giannakou E, Albracht K, Arampatzis A. Reproducibility of fascicle length and pennation angle of gastrocnemius medialis in human gait in vivo. *Gait Posture*. 2010; 31(1):73–7. Epub 2009/09/25. <https://doi.org/10.1016/j.gaitpost.2009.08.249> PMID: 19775893.
25. de Boer MD, Seynnes OR, di Prampero PE, Pisot R, Mekjavic IB, Biolo G, et al. Effect of 5 weeks horizontal bed rest on human muscle thickness and architecture of weight bearing and non-weight bearing

- muscles. *Eur J Appl Physiol*. 2008; 104(2):401–7. Epub 2008/03/06. <https://doi.org/10.1007/s00421-008-0703-0> PMID: 18320207.
26. Muramatsu T, Muraoka T, Takeshita D, Kawakami Y, Hirano Y, Fukunaga T. Mechanical properties of tendon and aponeurosis of human gastrocnemius muscle in vivo. *J Appl Physiol* (1985). 2001; 90(5):1671–8. Epub 2001/04/12. <https://doi.org/10.1152/jappl.2001.90.5.1671> PMID: 11299254.
 27. Raj IS, Bird SR, Shield AJ. Reliability of ultrasonographic measurement of the architecture of the vastus lateralis and gastrocnemius medialis muscles in older adults. *Clin Physiol Funct Imaging*. 2012; 32(1):65–70. <https://doi.org/10.1111/j.1475-097X.2011.01056.x> PMID: 22152081.
 28. Cronin NJ, Carty CP, Barrett RS, Lichtwark G. Automatic tracking of medial gastrocnemius fascicle length during human locomotion. *J Appl Physiol* (1985). 2011; 111(5):1491–6. Epub 2011/08/13. <https://doi.org/10.1152/jappphysiol.00530.2011> PMID: 21836045.
 29. Rana M, Hamarneh G, Wakeling JM. Automated tracking of muscle fascicle orientation in B-mode ultrasound images. *J Biomech*. 2009; 42(13):2068–73. Epub 2009/08/04. <https://doi.org/10.1016/j.jbiomech.2009.06.003> PMID: 19646699.
 30. Drazan JF, Hullfish TJ, Baxter JR. An automatic fascicle tracking algorithm quantifying gastrocnemius architecture during maximal effort contractions. *PeerJ*. 2019; 7:e7120. Epub 2019/07/16. <https://doi.org/10.7717/peerj.7120> PMID: 31304054; PubMed Central PMCID: PMC6611451.
 31. Zhou GQ, Chan P, Zheng YP. Automatic measurement of pennation angle and fascicle length of gastrocnemius muscles using real-time ultrasound imaging. *Ultrasonics*. 2015; 57:72–83. Epub 2014/12/04. <https://doi.org/10.1016/j.ultras.2014.10.020> PMID: 25465963.
 32. Konig N, Cassel M, Intziagianni K, Mayer F. Inter-rater reliability and measurement error of sonographic muscle architecture assessments. *J Ultrasound Med*. 2014; 33(5):769–77. Epub 2014/04/26. <https://doi.org/10.7863/ultra.33.5.769> PMID: 24764331.
 33. Klimstra M, Dowling J, Durkin JL, MacDonald M. The effect of ultrasound probe orientation on muscle architecture measurement. *J Electromyogr Kinesiol*. 2007; 17(4):504–14. Epub 2006/08/22. <https://doi.org/10.1016/j.jelekin.2006.04.011> PMID: 16919969.
 34. Loram ID, Maganaris CN, Lakie M. Use of ultrasound to make noninvasive in vivo measurement of continuous changes in human muscle contractile length. *J Appl Physiol* (1985). 2006; 100(4):1311–23. Epub 2005/12/13. <https://doi.org/10.1152/jappphysiol.01229.2005> PMID: 16339341.
 35. Morse CI, Thom JM, Birch KM, Narici MV. Changes in triceps surae muscle architecture with sarcopenia. *Acta Physiol Scand*. 2005; 183(3):291–8. Epub 2005/03/04. <https://doi.org/10.1111/j.1365-201X.2004.01404.x> PMID: 15743389.
 36. Oda T, Kanehisa H, Chino K, Kurihara T, Nagayoshi T, Fukunaga T, et al. In vivo behavior of muscle fascicles and tendinous tissues of human gastrocnemius and soleus muscles during twitch contraction. *J Electromyogr Kinesiol*. 2007; 17(5):587–95. Epub 2006/08/15. <https://doi.org/10.1016/j.jelekin.2006.04.013> PMID: 16904343.
 37. Benard MR, Becher JG, Harlaar J, Huijing PA, Jaspers RT. Anatomical information is needed in ultrasound imaging of muscle to avoid potentially substantial errors in measurement of muscle geometry. *Muscle Nerve*. 2009; 39(5):652–65. Epub 2009/03/18. <https://doi.org/10.1002/mus.21287> PMID: 19291798.
 38. Bolsterlee B, Gandevia SC, Herbert RD. Ultrasound imaging of the human medial gastrocnemius muscle: how to orient the transducer so that muscle fascicles lie in the image plane. *J Biomech*. 2016; 49(7):1002–8. <https://doi.org/10.1016/j.jbiomech.2016.02.014> PMID: 26905734.
 39. Franchi MV, Raiteri BJ, Longo S, Sinha S, Narici MV, Csapo R. Muscle Architecture Assessment: Strengths, Shortcomings and New Frontiers of in Vivo Imaging Techniques. *Ultrasound Med Biol*. 2018; 44(12):2492–504. Epub 2018/09/07. <https://doi.org/10.1016/j.ultrasmedbio.2018.07.010> PMID: 30185385.
 40. Sarto F, Sporri J, Fitze DP, Quinlan JI, Narici MV, Franchi MV. Implementing Ultrasound Imaging for the Assessment of Muscle and Tendon Properties in Elite Sports: Practical Aspects, Methodological Considerations and Future Directions. *Sports Med*. 2021; 51(6):1151–70. Epub 2021/03/09. <https://doi.org/10.1007/s40279-021-01436-7> PMID: 33683628; PubMed Central PMCID: PMC8124062.
 41. Noorkoiv M, Stavnsbo A, Aagaard P, Blazevich AJ. In vivo assessment of muscle fascicle length by extended field-of-view ultrasonography. *J Appl Physiol* (1985). 2010; 109(6):1974–9. Epub 2010/10/05. <https://doi.org/10.1152/jappphysiol.00657.2010> PMID: 20884841.
 42. Duclay J, Martin A, Duclay A, Cometti G, Pousson M. Behavior of fascicles and the myotendinous junction of human medial gastrocnemius following eccentric strength training. *Muscle Nerve*. 2009; 39(6):819–27. Epub 2009/03/21. <https://doi.org/10.1002/mus.21297> PMID: 19301364.
 43. Muramatsu T, Muraoka T, Kawakami Y, Shibayama A, Fukunaga T. In vivo determination of fascicle curvature in contracting human skeletal muscles. *J Appl Physiol* (1985). 2002; 92(1):129–34. Epub 2001/12/18. <https://doi.org/10.1152/jappl.2002.92.1.129> PMID: 11744651.

44. Rana M, Hamarneh G, Wakeling JM. 3D curvature of muscle fascicles in triceps surae. *J Appl Physiol* (1985). 2014; 117(11):1388–97. Epub 2014/10/18. <https://doi.org/10.1152/jappphysiol.00109.2013> PMID: 25324510; PubMed Central PMCID: PMC4254837.
45. Reeves ND, Narici MV, Maganaris CN. In vivo human muscle structure and function: adaptations to resistance training in old age. *Exp Physiol*. 2004; 89(6):675–89. Epub 2004/08/26. <https://doi.org/10.1113/expphysiol.2004.027797> PMID: 15328305.
46. Mohagheghi AA, Khan T, Meadows TH, Giannikas K, Baltzopoulos V, Maganaris CN. Differences in gastrocnemius muscle architecture between the paretic and non-paretic legs in children with hemiplegic cerebral palsy. *Clinical biomechanics*. 2007; 22(6):718–24. Epub 2007/05/04. <https://doi.org/10.1016/j.clinbiomech.2007.03.004> PMID: 17475377.
47. Koryak YA. Architectural and functional specifics of the human triceps surae muscle in vivo and its adaptation to microgravity. *J Appl Physiol* (1985). 2019; 126(4):880–93. Epub 2018/12/21. <https://doi.org/10.1152/jappphysiol.00634.2018> PMID: 30571290.
48. Eng JJ, Winter DA. Kinetic analysis of the lower limbs during walking: what information can be gained from a three-dimensional model? *J Biomech*. 1995; 28(6):753–8. Epub 1995/06/01. [https://doi.org/10.1016/0021-9290\(94\)00124-m](https://doi.org/10.1016/0021-9290(94)00124-m) PMID: 7601875.
49. Kurokawa S, Fukunaga T, Fukashiro S. Behavior of fascicles and tendinous structures of human gastrocnemius during vertical jumping. *J Appl Physiol* (1985). 2001; 90(4):1349–58. Epub 2001/03/15. <https://doi.org/10.1152/jappphysiol.2001.90.4.1349> PMID: 11247934.
50. Cenni F, Bar-On L, Schless SH, Kalkman B, Aertbelien E, Bruyninckx H, et al. Medial Gastrocnemius Muscle-Tendon Junction and Fascicle Lengthening across the Range of Motion Analyzed in 2-D and 3-D Ultrasound Images. *Ultrasound Med Biol*. 2018; 44(12):2505–18. Epub 2018/09/03. <https://doi.org/10.1016/j.ultrasmedbio.2018.07.012> PMID: 30172570.
51. Martin DC, Medri MK, Chow RS, Oxorn V, Leekam RN, Agur AM, et al. Comparing human skeletal muscle architectural parameters of cadavers with in vivo ultrasonographic measurements. *J Anat*. 2001; 199(Pt 4):429–34. <https://doi.org/10.1046/j.1469-7580.2001.19940429.x> PMID: 11693303; PubMed Central PMCID: PMC1468353.
52. Marfell-Jones MJ, Stewart A, De Ridder J. International standards for anthropometric assessment 2012.
53. Legerlotz K, Riley GP, Screen HR. Specimen dimensions influence the measurement of material properties in tendon fascicles. *J Biomech*. 2010; 43(12):2274–80. <https://doi.org/10.1016/j.jbiomech.2010.04.040> PMID: 20483410; PubMed Central PMCID: PMC2935962.
54. Baumbach SF, Brumann M, Binder J, Mutschler W, Regauer M, Polzer H. The influence of knee position on ankle dorsiflexion—a biometric study. *BMC musculoskeletal disorders*. 2014; 15:246. Epub 2014/07/24. <https://doi.org/10.1186/1471-2474-15-246> PMID: 25053374; PubMed Central PMCID: PMC4118219.
55. Cho KH, Lee HJ, Lee WH. Reliability of rehabilitative ultrasound imaging for the medial gastrocnemius muscle in poststroke patients. *Clin Physiol Funct Imaging*. 2014; 34(1):26–31. Epub 2013/06/14. <https://doi.org/10.1111/cpf.12060> PMID: 23758734.
56. Ihnatsenka B, Boezaart AP. Ultrasound: Basic understanding and learning the language. *Int J Shoulder Surg*. 2010; 4(3):55–62. Epub 2011/04/08. <https://doi.org/10.4103/0973-6042.76960> PMID: 21472065; PubMed Central PMCID: PMC3063344.
57. Knudsen L, Nawrotzki R, Schmiedl A, Muhlfeld C, Kruschinski C, Ochs M. Hands-on or no hands-on training in ultrasound imaging: A randomized trial to evaluate learning outcomes and speed of recall of topographic anatomy. *Anat Sci Educ*. 2018; 11(6):575–91. Epub 2018/04/24. <https://doi.org/10.1002/ase.1792> PMID: 29683560.
58. Huijing PA. Architecture of the Human Gastrocnemius Muscle and Some Functional Consequences. *Cells Tissues Organs*. 1985; 123(2):101–7. <https://doi.org/10.1159/000146047> PMID: 4061024
59. Ando R, Taniguchi K, Saito A, Fujimiya M, Katayose M, Akima H. Validity of fascicle length estimation in the vastus lateralis and vastus intermedius using ultrasonography. *J Electromyogr Kinesiol*. 2014; 24(2):214–20. <https://doi.org/10.1016/j.jelekin.2014.01.003> PMID: 24485560.
60. Seynnes OR, Cronin NJ. Simple Muscle Architecture Analysis (SMA): An ImageJ macro tool to automate measurements in B-mode ultrasound scans. *PLoS One*. 2020; 15(2):e0229034. Epub 2020/02/13. <https://doi.org/10.1371/journal.pone.0229034> PMID: 32049973; PubMed Central PMCID: PMC7015391.
61. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods*. 1996; 1(1):30–46. <https://doi.org/10.1037/1082-989x.1.1.30> WOS:A1996VV65700004.
62. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*. 2016; 15(2):155–63. Epub 2016/06/23. <https://doi.org/10.1016/j.jcm.2016.02.012> PMID: 27330520; PubMed Central PMCID: PMC4913118.

63. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res.* 1999; 8(2):135–60. Epub 1999/09/29. <https://doi.org/10.1177/096228029900800204> PMID: 10501650.
64. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. *Measurement in Medicine: A Practical Guide.* Cambridge: Cambridge University Press; 2011.
65. Portney LG. *Foundations of Clinical Research: Applications to Evidence-Based Practice.* FA Davis; 2020.
66. de Vet HC, Terwee CB, Ostelo RW, Beckerman H, Knol DL, Bouter LM. Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. *Health and quality of life outcomes.* 2006; 4:54. Epub 2006/08/24. <https://doi.org/10.1186/1477-7525-4-54> PMID: 16925807; PubMed Central PMCID: PMC1560110.
67. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res.* 2005; 19(1):231–40. Epub 2005/02/12. <https://doi.org/10.1519/15184.1> PMID: 15705040.
68. Haley SM, Fragala-Pinkham MA. Interpreting change scores of tests and measures used in physical therapy. *Phys Ther.* 2006; 86(5):735–43. Epub 2006/05/03. PMID: 16649896.
69. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis.* 1987; 40(2):171–8. Epub 1987/01/01. [https://doi.org/10.1016/0021-9681\(87\)90069-5](https://doi.org/10.1016/0021-9681(87)90069-5) PMID: 3818871.
70. Cohen J. *Statistical power analysis for the behavioral sciences.* Academic press; 2013.
71. Corp I. *IBM SPSS Statistics for Windows, Version 25.0.* Armonk, NY: IBM Corp; Released 2017.
72. Geremia JM, Baroni BM, Bini RR, Lanferdini FJ, de Lima AR, Herzog W, et al. Triceps Surae Muscle Architecture Adaptations to Eccentric Training. *Front Physiol.* 2019; 10:1456. Epub 2019/12/19. <https://doi.org/10.3389/fphys.2019.01456> PMID: 31849706; PubMed Central PMCID: PMC6901927.
73. Cenni F, Schless SH, Bar-On L, Molenaers G, Van Campenhout A, Aertbelien E, et al. Can in Vivo Medial Gastrocnemius Muscle-Tendon Unit Lengths be Reliably Estimated by Two Ultrasonography Methods? A Within-Session Analysis. *Ultrasound Med Biol.* 2018; 44(1):110–8. Epub 2017/11/11. <https://doi.org/10.1016/j.ultrasmedbio.2017.09.018> PMID: 29122315.
74. Turton P, Hay R, Taylor J, McPhee J, Welters I. Human limb skeletal muscle wasting and architectural remodeling during five to ten days intubation and ventilation in critical care—an observational study using ultrasound. *BMC Anesthesiol.* 2016; 16(1):119. Epub 2016/11/30. <https://doi.org/10.1186/s12871-016-0269-z> PMID: 27894277; PubMed Central PMCID: PMC5127036.
75. Morse CI, Smith J, Denny A, Tweedale J, Searle ND. Gastrocnemius medialis muscle architecture and physiological cross sectional area in adult males with Duchenne muscular dystrophy. *J Musculoskeletal Neuronal Interact.* 2015; 15(2):154–60. Epub 2015/06/03. PMID: 26032207; PubMed Central PMCID: PMC5133718.
76. Thom JM, Morse CI, Birch KM, Narici MV. Influence of muscle architecture on the torque and power-velocity characteristics of young and elderly men. *Eur J Appl Physiol.* 2007; 100(5):613–9. Epub 2007/05/29. <https://doi.org/10.1007/s00421-007-0481-0> PMID: 17530274.
77. Narici M, Cerretelli P. Changes in human muscle architecture in disuse-atrophy evaluated by ultrasound imaging. *J Gravit Physiol.* 1998; 5(1):P73–4. Epub 2001/09/07. PMID: 11542371.
78. Crouzier M, Lacourpaille L, Nordez A, Tucker K, Hug F. Neuromechanical coupling within the human triceps surae and its consequence on individual force-sharing strategies. *The Journal of Experimental Biology.* 2018; 221(21):jeb187260. <https://doi.org/10.1242/jeb.187260> PMID: 30237240
79. Kawakami Y, Abe T, Fukunaga T. Muscle-fiber pennation angles are greater in hypertrophied than in normal muscles. *J Appl Physiol (1985).* 1993; 74(6):2740–4. Epub 1993/06/01. <https://doi.org/10.1152/jappl.1993.74.6.2740> PMID: 8365975.
80. Strasser EM, Draskovits T, Praschak M, Quittan M, Graf A. Association between ultrasound measurements of muscle thickness, pennation angle, echogenicity and skeletal muscle strength in the elderly. *Age (Dordr).* 2013; 35(6):2377–88. Epub 2013/03/05. <https://doi.org/10.1007/s11357-013-9517-z> PMID: 23456136; PubMed Central PMCID: PMC3824993.
81. Fouré A, Nordez A, Cornu C. Effects of eccentric training on mechanical properties of the plantar flexor muscle-tendon complex. *J Appl Physiol (1985).* 2013; 114(5):523–37. Epub 2012/12/15. <https://doi.org/10.1152/jappphysiol.01313.2011> PMID: 23239873.
82. Bolsterlee B, Gandevia SC, Herbert RD. Effect of Transducer Orientation on Errors in Ultrasound Image-Based Measurements of Human Medial Gastrocnemius Muscle Fascicle Length and Pennation. *PLoS One.* 2016; 11(6):e0157273. Epub 2016/06/15. <https://doi.org/10.1371/journal.pone.0157273> PMID: 27294280; PubMed Central PMCID: PMC4905739.