Ultrasound in Med. & Biol., Vol. 00, No. 00, pp. 1−11, 2021

0301-5629/$ - see front matter

ELSEVIER

https://doi.org/10.1016/j.ultrasmedbio.2021.04.028

● *Original Contribution*

RELIABILITY OF PROCESSING 3-D FREEHAND ULTRASOUND DATA TO DEFINE MUSCLE VOLUME AND ECHO-INTENSITY IN PEDIATRIC LOWER LIMB MUSCLES WITH TYPICAL DEVELOPMENT OR WITH SPASTICITY

Britta Hanssen,[*,†,‡] Nathalie De Beukelaer,[*,†] Simon-Henri Schless,[*,†,§] Francesco Cenni,[†,¶,‖] Lynn Bar-On,[*,†,#] Nicky Peeters,[*,†,‡] Guy Molenaers,[**,††] Anja Van Campenhout,[**,††] Christine Van den Broeck,[‡] and Kaat Desloovere[*,†]

[*] Department of Rehabilitation Sciences, KU Leuven, Leuven, Belgium; [†] Clinical Motion Analysis Laboratory, University Hospitals Leuven, Pellenberg, Belgium; [‡] Department of Rehabilitation Sciences, Ghent University, Ghent, Belgium; [§] Motion Analysis and Biofeedback Laboratory, ALYN Paediatric and Rehabilitation Hospital, Jerusalem, Israel; [¶] Department of Mechanical Engineering, KU Leuven, Leuven, Belgium; [‖] Laboratory of Kinesiology Willy Taillard, Geneva University Hospitals and University of Geneva, Geneva, Switzerland; [#] Department of Rehabilitation Medicine, Amsterdam UMC, Amsterdam Movement Sciences, Amsterdam, The Netherlands; [**] Department of Development and Regeneration, KU Leuven, Leuven, Belgium; and [††] Orthopaedic Section, University Hospitals Leuven, Leuven, Belgium

(*Received* 22 *November* 2020; *revised* 23 *April* 2021; *in final from* 27 *April* 2021)

Abstract—This investigation assessed the processer reliability of estimating muscle volume and echo-intensity of the rectus femoris, tibialis anterior and semitendinosus. The muscles of 10 typically developing children (8.15 [1.40] y) and 15 children with spastic cerebral palsy (7.67 [3.80] y; Gross Motor Function Classification System I = 5, II = 5, III = 5) were scanned with 3-D freehand ultrasonography. For the intra-processer analysis, the intra-class correlations coefficients (ICCs) for muscle volume ranged from 0.943−0.997, with relative standard errors of measurement (SEM%) ranging from 1.24%−8.97%. For the inter-processer analysis, these values were 0.853 to 0.988 and 3.47% to 14.02%, respectively. Echo-intensity had ICCs >0.947 and relative SEMs <4% for both analyses. Muscle volume and echo-intensity can be reliably extracted for the rectus femoris, semitendinosus and tibialis anterior in typically developing children and children with cerebral palsy. The need for a single processer to analyze all data is dependent on the size of the expected changes or differences. (E-mail: Britta.hanssen@kuleuven.be)

Key Words: **3-D freehand ultrasonography, Reliability, Muscle volume, Echo-intensity, Cerebral palsy, Children.**

INTRODUCTION

Size and composition of skeletal muscles are important characteristics related to the force-producing capacity of the muscle (Fukunaga et al. 2001). Muscle tissue is very adaptive, and *in vivo* imaging can be used to evaluate adaptations to positive and negative stimuli such as treatment, training, aging and disuse (Fry et al. 2004; McNee et al. 2009; Barber et al. 2013). Therefore, evaluation of the morphological characteristics is relevant for both research and clinical practice.

Although magnetic resonance imaging (MRI) has been considered the gold standard for muscle volume (MV) evaluations, 3-D freehand ultrasonography (3-DfUS) has become an established alternative method (Cenni et al. 2016; Mozaffari and Lee 2017). Moreover, indirect information on intrinsic muscle composition can be derived from the echo intensity (EI) of ultrasound images, which has been related to the content of contractile and non-contractile tissue in the muscle (Pitcher et al. 2015; Young et al. 2015).

Three-dimensional fUS is an imaging technique combining conventional 2-D B-mode ultrasonography with a motion tracking system. The position and orientation of the US probe at every image can be combined to create 3-D data sets of anatomical volumes. By manually drawing transverse plane segmentations along the muscle border in the 2-D images and linearly interpolating these cross-sectional areas, a 3-D reconstruction of a muscle can be created from which the MV and average EI can be calculated (Cenni et al. 2016). It provides

Address correspondence to: Britta Hanssen, Weligerveld 1, 3212 Pellenberg, Belgium. E-mail: Britta.hanssen@kuleuven.be

good clinical utility considering the short acquisition times, low cost and ease of availability. Additionally, it offers possibilities to investigate muscle, tendon and fascicle lengths, as well as pennation angles, thus providing comprehensive information on muscle properties (Barber et al. 2011; Bland et al. 2011; Cenni et al. 2018a).

Three-dimensional fUS has been applied to assess differences in MV between typically developing (TD) children and children with the neurodevelopmental disorder cerebral palsy (CP), revealing lower MV in children with CP (Malaiya et al. 2007; Barber et al. 2011, 2016; Haberfehlner et al. 2016a; Schless et al. 2018b, 2019; Willerslev-Olsen et al. 2018). EI has also been compared between TD children and children with CP, with higher average EI values reported in the CP cohorts (Pitcher et al. 2015; Obst et al. 2017; Schless et al. 2018b, 2019). Additionally, 3-DfUS has been used to assess changes in MV in children with CP after treatment (Fry et al. 2007; McNee et al. 2009; Barber et al. 2013; Haberfehlner et al. 2018).

The accuracy of 3-DfUS has been assessed against ground truth objects (Kot et al. 2009; Cenni et al. 2016), cadavers (Weller et al. 2007; Haberfehlner et al. 2016b) and MRI assessments (Barber et al. 2009, 2018; Noorkoiv et al. 2019). Both Barber et al. (2018) and Noorkoiv et al. (2019) evaluated the accuracy of 3-DfUS in a CP population for the plantar flexor muscles and medial gastrocnemius (MG), respectively, revealing satisfactory results, that is, errors that were below the expected differences between TD children and children with CP. In addition to accuracy, it is important that the defined morphological muscle parameters are reliable, to provide researchers and clinicians with a good reference to evaluate the patient's muscle condition and the muscle's response to treatment (McMahon et al. 2016). For children with CP, these reliability data are crucial to interpret differences from a TD population, to evaluate individual muscle growth and to investigate changes after treatments such as botulinum toxin type A injections, surgery and resistance training.

Reliability is the extent to which measurements can be replicated (Daly and Bourke 2000). In the process of evaluating muscle morphological properties from 3-D fUS, the rater (*e.g.,* the researcher or clinician) has an influence during both the measurement part and the processing part. This results in two levels of reliability, acquirer and processer reliability, both of which can be assessed within an intra- and inter-rater perspective. The processer is a crucial factor in 3-D fUS evaluations, because US images must be investigated carefully to correctly identify the entire muscle and to properly outline the borders of the muscle's cross-sectional area. Consequently, standardized guidelines and training of processers are necessary. Although training experience of processers is sometimes briefly mentioned, specific guidelines on data processing have not been provided in previous publications. Furthermore, evaluation of processer reliability can help in the interpretation of inter- and intra-acquirer reliability, because part of the differences between repeated measures that are judged during the processing part can originate from the reliability limitations (Mozaffari and Lee 2017).

To date, the processer reliability of 3-D fUS has been assessed for volumes of ground truth objects, cadavers and lower limb muscles of adults and children. These previous findings indicated good to excellent reliability with intra-class correlation coefficients (ICCs) $\geq$0.920 (Barber et al. 2009, 2018; MacGillivray et al. 2009; Cenni et al. 2016, 2018b; Haberfehlner et al. 2016b). In children, only MVs of the plantar flexor muscles have been evaluated. Barber et al. (2018) studied the intra- and inter-processer reliability of the soleus, medial and lateral gastrocnemius in children with CP and obtained ICCs $\geq$0.967. Likewise, Cenni et al. (2018b) investigated the inter-processer reliability for evaluating MV of the MG in both children with CP and TD children, with ICCs for MV $>$0.994 in both cohorts. This investigation also studied the EI of the MG and obtained ICCs of 0.979 in TD children and 0.886 in children with CP.

Three-dimensional fUS reliability studies in TD children and children with CP have been limited to the plantar flexor muscles. Yet, researchers have already investigated morphological differences between populations, associations with functional parameters and changes after interventions for other lower limb muscles such as the rectus femoris (RF), tibialis anterior (TA) and semitendinosus (ST) (Bland et al. 2011; Williams et al. 2012; Moreau et al. 2013; Lee et al. 2015; Handsfield et al. 2016; Haberfehlner et al. 2018; Schless et al. 2019). Therefore, the aim of this investigation was to determine the within-session intra- and inter-processer reliability for estimating MV and EI of three lower limb muscles (RF, TA and ST) in TD children and children with CP. Secondarily we compared the magnitude of the processing errors with the differences in MV and EI between the cohorts. We hypothesized that the intra-processer condition is more reliable than the inter-processer condition and that there are differences in reliability between muscles because of their difference in muscle architecture.

METHODS

Participants

Earlier published effect sizes for EI and normalized MV of the MG and TA muscles indicated that at least 5 TD children and 5 children with spastic CP (SCP) would be sufficient to find significant differences between TD and SCP (Schless et al., 2019). To also ensure sufficiently large samples for the reliability assessments, 10 TD children with no known neurological or orthopedic problems were recruited via hospital co-workers and students. Additionally, 15 children with SCP were recruited via the Clinical Motion Analysis Laboratory of the University Hospital in Pellenberg (Walter et al. 1998). Children were approached if a diagnosis of SCP was confirmed by the medical record and functionality ranged between levels I and III on the Gross Motor Function Classification System (GMFCS) (Palisano et al. 2008). Lower limb surgery within 2 y and botulinum toxin A treatment within 6 mo before the assessments were used as exclusion criteria. All data were collected as part of ongoing projects that were approved by the local ethics committee of the University Hospitals of Leuven. Written informed consent was acquired from all parents of the participants as well as from participants 12 y and older.

Acquisition

A previously described 3-D fUS technique was applied to perform single-sweep assessments of the RF, TA and ST muscles (Cenni et al. 2016). A B-mode US device was used with a 5.9-cm-wide linear transducer with a frequency of 10 MHz (HL9.0/60/128Z, Echoblaster 128 Ext-1Z system, Telemed, Vilnius, Lithuania). US images were acquired at a frequency of 30 Hz while US settings were kept constant for all assessments within the cohort of TD children and within the cohort of SCP children. Four optical markers were rigidly mounted on the US transducer and tracked with a portable motion tracking system with three fixed optical cameras, a sampling rate of 120 Hz and a spatial resolution of 1 mm (Optitrack V120:Trio, NaturalPoint, Corvallis, OR, USA). Synchronized acquisition was established by applying a signal trigger with the US device as master. The respective retail software was used for acquiring all data. Because US settings could differ between cohorts (ranges of settings—frequency: 8−10 MHz, depth: 50 mm, focus: 12−20 mm, gain: 52%−58%, dynamic range: 68−80 dB, time-gain compensation: neutral [50%], power: 100%), the acquired US images were converted to common reference settings. Hereto, a cross-calibration was applied to US images obtained through standardized measurements with each used set of settings, as based on a technique described before (Pillen et al. 2009), and a conversion equation was calculated to the reference settings. It should be noted that the conversions did not influence the repeatability of EI, which is the primary focus of the current study, but caused a change in the absolute values, allowing comparison between groups, as a secondary focus of the study.

A random leg was chosen for the TD children, while the most affected leg according to their most recent clinical examination was measured in the participants with SCP. The RF and TA were assessed in supine position and the ST in prone position, all with a triangular cushion placed under the shank providing approximately 25° of knee flexion and an unconstrained ankle position. Large amounts of acoustic transmission gel were applied during the acquisition while the US transducer was moved from the proximal to the distal aspect over the targeted muscle at a constant velocity. US images were acquired in a transverse orientation while the transducer was held perpendicular to the deep aponeurosis of the muscle. All acquisitions were performed by one of two experienced researchers (S.H.S. or F.C.) and were repeated in case of suspected movement of the leg or muscle contraction.

Data processing

Three-dimensional data sets were created using an open-source software library in Python (Cenni et al. 2016). With a custom workflow in MeVisLab (https://www.mevislab.de), equally spaced manual segmentations were performed, outlining the inside of the muscle border in the transverse plane of the 3-D reconstructed anatomical volume (Fig. 1). The number of segmented images was expressed as a percentage of all acquired images from origin to muscle tendon junction. The percentage of segmented images was not fixed and could be freely chosen for each muscle by the processers. A linear interpolation between the outlined borders was applied to compute the MV (in mL). To compare MV between cohorts, the data were normalized to body weight (nMV in mL/kg). The mean EI (expressed on a gray scale of 256 values, arbitrary units [AU]) was computed within the interpolated reconstruction of the whole muscle. The muscles were segmented from the origin to the muscle tendon junction. Next to total volume, the ST was also evaluated in two parts: the proximal part (ST_{Prox}) from the ischial tuberosity down to the end of the internal aponeurosis, and the distal part (ST_{Dist}) from the start of the internal aponeurosis to the muscle tendon junction (Haberfehlner et al. 2016b).

Study design

Two processers (processer 1 [B.H.] and processor 2 [N.D.B.]) independently analyzed the same 3-D data sets of three muscles for all 25 participants. One processer repeated the analyses with at least a 2-wk period between (analyses 1A and 1B). Before the processing procedure all files were anonymized such that processers were unaware of the participants' characteristics. Before the

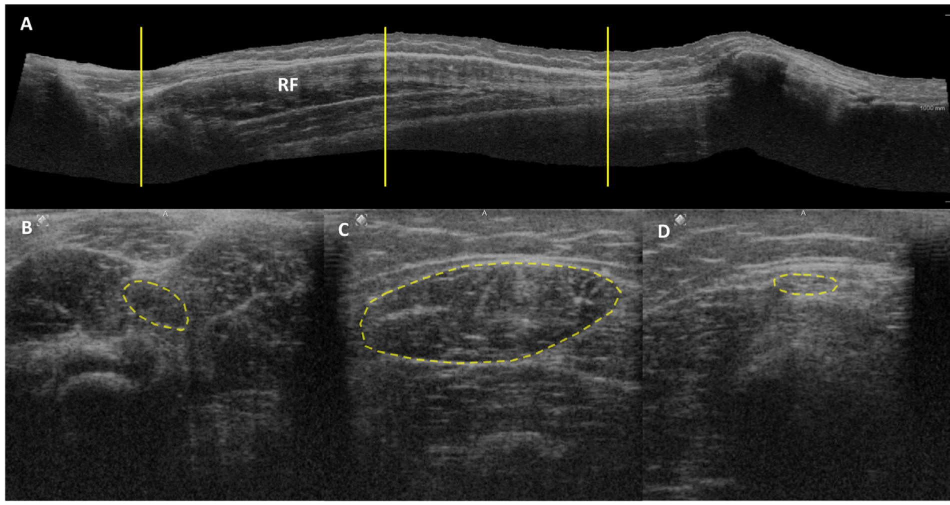

Fig. 1. Example of a 3-D data set of the rectus femoris (RF) with outlined muscle borders in three transverse plane images. (A) Three-dimensional data set of a scan of the rectus femoris with the *vertical lines* corresponding to the three transverse plane images shown below. (B) Transverse plane image around the origin. (C) Transverse plane image around the middle of the muscle belly. (D) Transverse plane image around the muscle tendon junction.

analyses, a 15-h training was given by one highly experienced processer (S.H.S.); it focused on general US background, recognition of bony landmarks, identification of proximal origin of the muscle and distal muscle−tendon transition and familiarization with (changes in) muscle shape (Silvestri et al. 2015). This training resulted in a set of general and muscle-specific guidelines. Next to 15 h of combined training, each processer had performed at least 50 3-D fUS measurements of the concerned muscles. Finally, before onset of the current repeatability study, practice data sets of each processer were evaluated by the experienced processer, and individual feedback was provided. An overview of each part of the training process (theory, guidelines and practice) is presented in Table 1, whereas specific guidelines for each muscle are provided in the Supplementary Data (Supplementary Figs. S1−S6, online only).

Statistical analyses

Intra- and inter-processer relative reliability was defined by means of the intra-class correlation coefficient (ICC 2,1) with 95% confidence intervals based on a single-rater, absolute-agreement, two-way random-effects model (Weir et al. 2005; Koo and Li 2016). Absolute reliability was defined by the standard error of measurement (SEM) as calculated from the square root of the mean square error from a one-way repeated-measures analysis of variance and was also expressed as a percentage of the mean of the two processers (SEM%). ICC scores <0.5 indicate poor reliability, while values between 0.5 and 0.75, between 0.75 and 0.9 and >0.9 indicate moderate, good and excellent reliability, respectively (Koo and Li 2016). Differences between cohorts and between processers were assessed with the Mann−Whitney *U*-test. A Bonferroni correction was applied to correct for multiple comparisons (p = 0.01). Finally, Scatter and Bland−Altman plots were created and independently checked by two assessors to determine any systematic bias.

RESULTS

All collected 3-D data sets for the RF, TA and ST were segmented for and included in the analysis. Descriptive statistics and comparisons between the cohorts for nMVs, EI values and percentages of segmented images can be found in Table 2. Age and weight did not differ significantly between the cohorts. There were significant differences for nMV of the RF and TA, with lower nMV in the SCP cohort. EI differed significantly for the RF and was close to significance for the TA (p = 0.031), with higher EI in the SCP cohort. The proportion of segmented images ranged from 3.47%−6.93% within the TD cohort and from 5.41%−9.32% in the SCP cohort, with significant differences found between cohorts for RF, TA and ST_{Tot}.

For the intra-processer analysis, the ICCs for MVs ranged from 0.943−0.997 and relative SEM scores ranged from 1.24%−8.97%. For EI, ICCs ranged from 0.952−0.996 with SEM% between 0.86% and 3.11%. On the basis of the ICC classification, this indicates excellent reliability for both parameters. For the inter-processer analysis, the ICCs for MVs ranged from 0.853−0.988, and relative SEM scores ranged from 3.47%−14.02%. Both the lowest ICC value and highest SEM% value were for the ST muscle. EI had ICCs

Table 1. Overview of the training protocol*

1. **Theoretical training: general knowledge of ultrasound and anatomy**
The first phase included study of the following basic knowledge packages:

- General background on ultrasound principles and technical specifications
- Anatomy of the muscle of interest and surrounding muscles
- Recognition of bony landmarks within the 3-D data sets for orientation
- Recognition of proximal origin of the muscle and distal muscle−tendon transition
- Familiarization with (changes in) muscle shape and aponeuroses

This educational training phase was supported by relevant scientific papers (Barber et al. 2009, 2018; Cenni et al. 2016, 2018b; Haberfehlner et al. 2016b) and handbooks (especially Silvestri et al. 2015) and resulted in a 3-D focused ultrasound manual with example data developed by the local research team.

2. **Familiarization with study specific guidelines and agreements**
Before the start of routine processing, guidelines or agreements on the following aspects should be made:

- The placement of the segmentation with respect to the aponeurosis surrounding the muscle needs to be standardized as inside, outside or on top of the aponeurosis.
- The distribution of the segmentations needs to be standardized. This could be an equal distribution throughout the muscle belly or a distribution that varies depending on the shape of the muscle (*e.g.,* with a larger number of segmentations for muscle regions that are characterized by an irregular shape and a lower number of segmentations for muscle regions that have a quite constant shape).
- The percentage of images to be segmented needs to be standardized per muscle or per study. A previous validity study recommended aiming for 10% of the images (Cenni et al. 2018b).
- Because of the rather noisy nature of ultrasound muscle images, specific guidelines on quality checks and data curation need to be defined. This involves specific agreements for inclusion or exclusion of data sets per parameter, based on the data quality. Insufficient quality of the muscle images that may lead to data exclusion for specific outcome parameters include:
    - The muscle width may exceed the width of the ultrasound probe. In that case the muscle border may be partly missed. The muscle border may also be difficult to detect when it is too deep. Such artifacts lead to data exclusion for MV and EI.
    - When the proximal or distal (bony) landmarks are not detectable in the muscle images, the data are excluded for MV and EI.
    - Insufficient contact between skin and probe, causing "black gaps" in the images. Such image artifacts do not necessarily affect the MV parameters, but lead to data exclusion for the EI parameter. It should be noted that such artifacts may also influence the MV if they appear in a substantial part of the reconstruction with negative impact on the muscle border recognition.
    - When the overall quality of the muscle images is too low, because of other measurement errors or because the degree of disease progression causes a lack of contrast between the muscle mass and aponeuroses, the majority of the muscle images may not be properly segmented and the data are excluded for MV and EI.

3. **Practical training: hands-on experience of ultrasound acquisition and processing**
Before the start of routine processing of 3-D fUS images, the involved processer should have had sufficient hands-on experience, for both data acquisition and data processing. It is recommended that the processor starts with healthy muscles before pathological muscles. The minimum number for each part is dependent on the concerned muscle(s), the prior knowledge of the assessor and his or her learning pace. The following steps are recommended:

- Observing data acquisitions on the muscles of interest in healthy and clinical populations by an experienced assessor.
- Performing data acquisitions on the muscles of interest in both typically developing and clinical populations, with a special focus on the points mentioned in the training phase (*i.e.,* shape of the muscle and aponeuroses).
- During practice data acquisitions, exploration of the anatomy around the specific muscle and asking participants for a contraction during scanning are recommended, as these help to identify the muscle's border with surrounding muscles and the shape in rest versus contraction (*e.g.,* dorsiflexion vs. toe extension for the tibialis anterior).
- Processing of practice data sets from both typically developing and clinical populations, provided with guidance and feedback from the experienced processer. This can be individual or discussed during the training workshops. The repeatability data of the current study can be used as a reference to judge the consistency of the novel processer(s).

EI = echo-intensity; MV = muscle volume.
* The protocol includes three parts, which may partly overlap during the training process.

ranging from 0.947−0.990, with the SEM% between 1.47% and 3.89%. Therefore, inter-processer reliability can be classified as good to excellent. Detailed information on ICCs and SEMs is given in Table 3. There was no clear tendency of muscles of children with SCP to have lower reliability indices than TD children, except for the ST.

The percentage of segmented images between processers were similar except for the ST_{prox} and ST_{dist}, where it differed between processers 1 and 2 for the TD cohort. No statistical differences between processers were found for MV or EI values. However, careful study of the Bland−Altman plots with limits of agreement suggested a tendency toward a difference between the two processers for ST_{dist} and ST_{tot} MVs (Supplementary Data, Appendix 2, online only).

DISCUSSION AND SUMMARY

The current investigation revealed reliable extraction of MV and EI from 3-D fUS reconstructions of RF, TA and ST muscles in cohorts of children with typical development and those with SCP. This was found for both intra-processor and an inter-processer conditions, where the intra-processer yielded slightly better reliability results.

While ICC values suggested good to excellent reliability, it should be noted that the lower limit of the confidence intervals indicates that in the inter-processer analyses, this ICC value is prone to variation for the TA and ST muscles because lower limits of confidence intervals ranged between 0.296 and 0.947, ranging from poor to excellent reliability on the lower end of the interval.

Table 2. Descriptive information and comparison between the typically developing and spastic cerebral palsy cohorts

	Typically developing (N = 10)		Spastic cerebral palsy (N = 15)		p Value (Mann−Whitney U-test)
	Median	(IQR)	Median	(IQR)	
Age (y)	8.15	(1.40)	7.67	(3.80)	NS
Weight (kg)	26.20	(6.60)	20.50	(7.80)	NS
Normalized muscle volume (mL/kg)					
Rectus femoris	2.66	(0.72)	1.90	(0.55)	0.008
Tibialis anterior	1.36	(0.34)	0.84	(0.19)	<0.001
Semitendinosus					
Total	1.79	(0.45)	1.90	(0.24)	NS
Proximal	0.90	(0.37)	0.93	(0.30)	NS
Distal	0.85	(0.21)	0.92	(0.27)	NS
Echo-intensity (AU)					
Rectus femoris	80.80	(7.82)	94.63	(22.25)	0.008
Tibialis anterior	95.15	(20.48)	112.16	(19.02)	NS
Semitendinosus					
Total	83.82	(13.19)	91.18	12.90	NS
Proximal	74.71	(17.27)	81.09	(13.51)	NS
Distal	98.39	(12.51)	100.22	(14.76)	NS
Segmented images (%)					
Rectus femoris	3.81	(0.47)	5.63	(1.81)	<0.001
Tibialis anterior	3.47	(0.55)	5.41	(1.56)	<0.001
Semitendinosus					
Total	5.85	(1.42)	9.32	(4.27)	<0.001
Proximal	6.93	(2.41)	7.76	(2.64)	NS
Distal	5.63	(2.26)	6.92	(2.76)	NS

AU = arbitrary units; IQR = interquartile range; NS = not significant.

Moreover, although statistical tests revealed no significant differences between processers, some systematic bias was visible in the Bland−Altman plots (Supplementary Data, Appendix 2, online only). This furthermore emphasizes the importance of sufficient training and clear guidelines among processers and single processer analyses to avoid the influence of the rater in cases where smaller differences are expected.

Previous reliability studies of 3-D fUS in children focused only on the plantar flexors. Barber et al. (2018) reported ICCs ranging from 0.983−0.966 for intra-processer analyses on the volumes of three plantar flexor muscles, with SEMs (derived from reported MDC values) of approximately 1.12 and 1.56 mL in the intra-processor and inter-processer conditions, respectively. For the inter-processer analyses, the ICCs ranged from 0.967−0.993 (Barber et al. 2018). Likewise, Cenni et al. (2018a) reported inter-processer results for the MG in TD children and children with SCP. The analyses of MV resulted in ICCs of 0.994 in the TD and 0.997 in the SCP cohort, with SEMs of 1.2 and 1.4 mL, respectively. These results are similar to the MV results of the intra-processer analyses in the current study, but better than the present inter-processer results. For EI, the ICC was 0.979 with an SEM of 1.2 AU in the TD cohort and an ICC of 0.886 with an SEM of 2.1 AU in the SCP cohort (Cenni et al. 2018b). In comparison, the ICCs for EI are similar to the ICCs in the current study, for both intra- and inter-processer analysis, as well as for both cohorts. The SEMs are also similar for the RF, TA and ST_{Tot}; however, they are a little higher in the current study for the separate parts of the ST.

For MV, the ICC and SEM results of the intra-processer analyses were always better than the results of the inter-processer analyses. For EI, the ICCs and SEMs revealed similar reliability for intra- and inter-processer analyses, with some exceptions for the ST. Comparison of the three analyzed muscles revealed that the RF had the highest ICCs for MV, followed by TA and then ST. Although the absolute SEMs were comparable for the RF and TA, the relative SEMs were higher for the TA because of its smaller volume. For EI, there was little difference between the RF and TA with respect to both ICCs and SEMs. The ST frequently had higher relative and absolute SEM values for both MV and EI. As had been reported before by Haberfehlner et al., 2016b, the ST is hard to visualize, especially the proximal part, because of the deeper anatomical position, leading to reduced US image quality from beam attenuation. An example of a more challenging ST has been added in Supplementary Figure S7 (online only). It should be noted that the visualization in 3-D aids in identification of the muscle. However, our results have similar reliability for the proximal and distal parts, whereas the MV of the combined proximal and distal parts had better results. This suggests that defining the beginning and end of the internal aponeurosis may have a crucial impact on reliability, implying the need for a high level of anatomical

Table 3. Absolute and relative reliability results for the inter- and intra-processer analyses of muscle volume and echo-intensity

Muscle volume	Intra-processer				Inter-processer			
	Mean MV_{1A+B}	ICC (CI)	SEM		Mean MV_{1A+2}	ICC (CI)	SEM	
			Absolute (mL)	Relative (%)			Absolute (mL)	Relative (%)
Rectus femoris								
TD	70.26	0.997 (0.990−0.999)	0.87	1.24	68.59	0.967 (0.740−0.993)	2.38	3.47
SCP	52.46	0.996 (0.987−0.999)	1.30	2.48	51.14	0.988 (0.947−0.996)	1.90	3.72
Tibialis anterior								
TD	36.38	0.989 (0.957−0.997)	0.87	2.39	36.22	0.936 (0.775−0.983)	2.54	7.01
SCP	23.59	0.985 (0.947−0.995)	1.00	4.23	22.60	0.924 (0.401−0.981)	1.46	6.45
Semitendinosus Total								
TD	50.71	0.982 (0.929−0.996)	1.84	3.62	49.05	0.948 (0.663−0.988)	3.15	6.42
SCP	50.33	0.983 (0.951-0.994)	2.67	5.30	47.76	0.942 (0.601−0.985)	4.20	8.80
Proximal							3.65	
TD	25.94	0.943 (0.797−0.986)	2.03	7.84	26.01	0.915 (0.695−0.978)		14.02
SCP	24.16	0.963 (0.892−0.987)	1.69	6.98	23.60	0.900 (0.727−0.965)	3.17	13.44
Distal								
TD	24.08	0.968 (0.885−0.992)	1.07	4.43	22.69	0.883 (0.319−0.974)	1.64	7.25
SCP	25.21	0.961 (0.887−0.987)	2.26	8.97	23.27	0.853 (0.296−0.959)	2.98	12.82

Echo-intensity	Intra-processer				Inter-processer			
	Mean EI_{1A+B}	ICC (CI)	SEM		Mean EI_{1A+2}	ICC (CI)	SEM	
			Absolute (AU)	Relative (%)			Absolute (AU)	Relative (%)
Rectus femoris								
TD	84.23	0.994 (0.975−0.998)	1.12	1.33	83.95	0.987 (0.950−0.997)	1.62	1.93
SCP	96.43	0.996 (0.988−0.999)	0.87	0.90	96.41	0.989 (0.968−0.996)	1.44	1.49
Tibialis anterior								
TD	100.33	0.991 (0.945−0.998)	1.11	1.10	99.61	0.990 (0.962−0.998)	1.48	1.48
SCP	112.00	0.989 (0.969−0.996)	1.25	1.11	111.47	0.974 (0.911−0.992)	1.68	1.51
Semitendinosus Total								
TD	87.24	0.992 (0.970−0.998)	1.33	1.53	86.67	0.976 (0.887−0.994)	1.99	2.30
SCP	91.50	0.968 (0.907−0.989)	1.89	2.06	90.74	0.966 (0.895−0.989)	1.75	1.93
Proximal								
TD	78.11	0.969 (0.876−0.992)	2.43	3.11	78.05	0.956 (0.840−0.989)	3.04	3.89
SCP	83.16	0.952 (0.865−0.984)	2.48	2.98	82.53	0.947 (0.847−0.982)	2.38	2.89
Distal								
TD	98.58	0.996 (0.981−0.999)	0.84	0.86	98.51	0.988 (0.955−0.997)	1.58	1.60
SCP	101.75	0.979 (0.939−0.993)	1.40	1.37	101.31	0.981 (0.945−0.994)	1.49	1.47

AU = arbitrary units; CI = confidence interval; EI = echo-intensity; ICC: intra-class correlation coefficient; MV = muscle volume; SCP = spastic cerebral palsy cohort; SEM = standard error of measurement; TD = typically developing cohort.

knowledge and experience necessary to be able to process 3-D fUS data. In comparison to previously reported experience levels of >1000 3-D fUS muscle scans (Barber et al. 2018), the experience level in this investigation was lower, yet still led to good to excellent reliability results. To facilitate clinical implementation of 3-D fUS, clear guidelines on prior knowledge and level of experience, as well as educational material and practice data sets, should be provided. An overview of the general training process is presented in Table 1, whereas the specific processing guidelines for each muscle are provided in Appendix 1 (Supplementary Data, online only).

There was little difference in reliability results between the cohorts of TD children and children with SCP, despite the previously reported differences in muscle morphological properties between these groups (Noble et al. 2014a, 2014b; Barber et al. 2016; Schless et al. 2018a, 2019). These differences in morphological muscle properties were also observed in the current study (Table 2). Despite the overall similarity in reliability, muscles of children with SCP were often experienced as more difficult to segment compared with TD muscles, especially as GMFCS levels increased. Yet, this has partially been compensated by segmenting a larger percentage of images in the SCP cohort.

This investigation found significant differences in nMV between cohorts of TD and SCP children for the RF and TA. Absolute differences of 18.1 mL for RF, indicating a deficit of 26%, and 13.4 mL for TA, indicating a deficit of 34%, were found. The EI values also significantly differed for the RF and were close to significance for TA, with absolute differences of 17.0 and 13.8 AU, respectively. There were no significant differences in MV or EI for the ST muscle. The differences between the TD and SCP cohorts for MV and EI of the RF and TA are comparable to differences reported in an investigation in which the relationship of MG and TA morphology with gait characteristics was assessed (Schless et al. 2019). The SEM values found in the current investigation ranged from 0.87−4.20 mL for MV and from 0.84−3.04 AU for EI while differences between TD and CP for TA and RF were >13 mL and 13 AU, respectively. The literature also presents absolute differences <10 mL between MG MV for children with GMFCS level I versus level II (Schless et al. 2018a). No significant differences were found for the ST in the current study; however, previous investigations in older children and children more severely affected with CP found differences in comparison to a TD population of more than 36 mL (62%) (Haberfehlner et al. 2016a) and decreases of 15 mL (44%) after medial hamstring lengthening (Haberfehlner et al. 2018). Therefore, our findings suggest that for cross-sectional analyses and comparisons with sufficiently large expected differences, the use of multiple assessors is justified. However, for interventions and follow-up studies with changes expected to be smaller than the reported SEMs from the inter-processer analyses, it is important to have a single processer analyzing the results.

There are several limitations to this study. First, this study did not report concurrent or criterion validity of the 3-D fUS technique for the investigated muscles. Although validity has previously been evaluated on cadavers and by means of MRI studies, this was only performed for a CP cohort for the plantar flexor muscles and with a different US protocol (Barber et al. 2018). However, the validity of the current 3-D fUS technique and protocol for extracting MV had been evaluated on ground truth objects, resulting in an accuracy of 3 mL (Cenni et al. 2016). Second, the current study reports only on processing repeatability. These analyses do not include the influence of repeated data collections at multiple time points. Therefore, a comprehensive reliability evaluation incorporating inter- and intra-acquirer as well as inter-session measurements should be carried out. Moreover, because the intra-processer part was based only on one observer and the inter-processer only on two processers, the results are not generalizable to larger groups of processers. Researchers and clinicians willing to implement this technique should perform their own intra- or inter-processer reliability analyses to define their level of precision. The training protocol described in Table 1 can thereby be used as a road map. Third, the current investigation included only relatively young children (median: 7.75 y, range: 5.1−15.9 y), which did not require multiple parallel sweeps within the same acquisition to visualize the whole muscle. Such multiple parallel sweeps are needed when the muscle width is larger than the width of the probe. Because previous investigations experienced inconsistencies in reconstructions with multiple parallel sweeps (Barber et al. 2018), these should be further evaluated in future studies. Finally, the segmentation procedure in the current study was manual, which is time consuming, especially for applications in clinical practice. In this investigation, it was found that segmenting 3% to 10% of the images still resulted in reliable results. However, it is important to make prior agreements on the percentage of segmented images when comparing cohorts or in follow-up studies to avoid the influence of over- or underestimating because of differences in the number of segmented images. Use of a semi-automatic or automatic segmentation procedure to define the outline of the muscle would be advantageous

(Cenni et al. 2018b). Similarly, because ultrasound settings influence the absolute value of EI, it is important to have site-specific reference data if comparisons between cohorts are desired.

Implications and recommendations

Although this investigation focused on a specific population (SCP) and three specific lower limb muscles, it highlighted some relevant topics for future research and for clinical use of 3-D fUS. First, a standardized comprehensive training protocol for US processing is considered crucial. Such a training protocol includes three parts: (1) the acquisition of crucial general theoretical background knowledge of US and anatomy, including muscle-specific indications on the expected shape of the muscle, relevant bony landmarks and appearance of surrounding muscle in the images; (2) familiarization with specific processing guidelines; and (3) the practical hands-on experience of ultrasound processing on practice databases. The training protocol that was developed for the current study is described in Table 1. In practice, the training of US processing is ideally combined with practical training on data acquisition because it supports the familiarization process with the muscle images. Moreover, although the data collection technique is not within the scope of the current study, a high-level muscle imaging acquisition is a prerequisite to achieve an overall good repeatability for the 3-D fUS technique. Some guidelines are study specific, such as agreements on the target number and distribution of the segmentations that need to be performed per muscle.

The current study created novel information that is relevant for different applications of the 3-D fUS technique. First, the data and training protocol provide a framework to study muscle morphology with 3-D fUS in several populations, including other patient groups that are characterized by spasticity, such as hereditary spastic paraplegia (De Beukelaer et al. 2021) and stroke (Schillebeeckx et al. 2020); pediatric populations with other underlying causes of disability, such as neuromuscular disorders or clubfoot; athletes with muscular injuries; and the elderly to study the effects of sarcopenia (Watanabe et al. 2013). In the view of future use of the current study findings for other populations, it should be noted that when the 3-D fUS technique is applied for larger muscles, for example, in adults or dystrophic pseudo-hypertrophic muscles, a good calibration for multiple parallel sweep data collection is a prerequisite (Cenni et al. 2019), which may also require additional processing guidelines.

Second, the reliability results of the current investigation provide relevant reference values against which to judge differences in muscle morphology parameters between clinically relevant subgroups of the CP population (*e.g.,* based on GMFCS level or age) in cross-sectional studies as well as to interpret changes in muscle morphology as an effect of interventions and/or as part of natural development in longitudinal follow-up studies. In that respect, our findings suggested that multiple processers can be involved in the processing phase when previous studies suggest large between-group differences, such as when comparing CP data with TD data. However, when more subtle differences may be expected, for example, after a training intervention or in a longitudinal study evaluating growth or deterioration during aging, single-processer analyses are recommended. Improved accuracy can be achieved if repeated acquisitions are processed or if multiple processing repetitions of the same acquisition are made, either by the same processer or by multiple processers, where the average of the multiple repetitions can be used as a final outcome (Barber et al. 2018).

For EI, current research has focused largely on cross-sectional comparisons or associations of muscle quality with, for example, muscle strength (Watanabe et al. 2013). If EI is of interest in follow-up studies after interventions or in longitudinal studies, we believe there is a greater influence of the acquisition technique than of the processing technique. A well-standardized measurement protocol with constant US settings is advised. This can be combined with 2-D acquisitions at multiple areas throughout the muscle, complementary to the 3-D reconstruction.

CONCLUSIONS

In young TD children and children with SCP, MV and EI can be reliably extracted for the RF, TA and ST muscles with similar reliability indices for both cohorts. These findings complement the previously reported intra- and inter-processer reliability of the MG. The intra-processer analyses resulted in slightly better results than the inter-processer analyses, indicating the importance of qualitative processer training and well-defined guidelines, as well as the need for single processer analyses in case of smaller expected changes or differences. Yet, these reliability results and the high clinical utility of 3-D fUS open up possibilities for both researchers and clinicians to assess differences between cohorts and monitor changes in relation to growth and interventions.

Acknowledgments—This work was supported by TBM Grant TAMTA-T005416N from the Research Foundation Flanders (FWO), Belgium, awarded to K.D. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the article.

Conflict of interest disclosure—The authors declare that they have no affiliations with or involvement in any organization with any financial or non-financial interest in the subject matter and materials discussed in this article.

SUPPLEMENTARY MATERIALS

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ultrasmedbio.2021.04.028.

REFERENCES

Barber L, Barrett R, Lichtwark G. Validation of a freehand 3D ultrasound system for morphological measures of the medial gastrocnemius muscle. J Biomech 2009;42:1313–1319.

Barber L, Hastings-Ison T, Baker R, Barrett R, Lichtwark G. Medial gastrocnemius muscle volume and fascicle length in children aged 2 to 5 years with cerebral palsy. Dev Med Child Neurol 2011;53:543–548.

Barber L, Hastings-Ison T, Baker R, Kerr Graham H, Barrett R, Lichtwark G. The effects of botulinum toxin injection frequency on calf muscle growth in young children with spastic cerebral palsy: A 12-month prospective study. J Child Orthop 2013;7:425–433.

Barber L, Read F, Lovatt Stern J, Lichtwark G, Boyd RN. Medial gastrocnemius muscle volume in ambulant children with unilateral and bilateral cerebral palsy aged 2 to 9 years. Dev Med Child Neurol 2016;53:543–548.

Barber L, Alexander C, Shipman P, Boyd R, Reid S, Elliott C. Validity and reliability of a freehand 3D ultrasound system for the determination of triceps surae muscle volume in children with cerebral palsy. J Anat 2018;234:384–391.

Bland DC, Prosser LA, Bellini LA, Alter KE, Damiano DL. Tibialis anterior architecture, strength and gait in individuals with cerebral palsy. Muscle Nerve 2011;44:509–517.

Cenni F, Monari D, Desloovere K, Erwin A, Schless S, Bruyninckx H. The reliability and validity of a clinical 3D freehand ultrasound system. Comput Methods Programs Biomed 2016;136:179–187.

Cenni F, Bar-On L, Schless SH, Kalkman B, Aertbelien E, Bruyninckx H, Desloovere K. 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 2018a;44:2505–2518.

Cenni F, Schless S, Bar-on L, Aertbeliën E, Bruyninckx H, Hanssen B, Desloovere K. Reliability of a clinical 3D freehand ultrasound technique: Analyses on healthy and pathological muscles. Comput Methods Programs Biomed 2018b;156:97–103.

Cenni F, Monari D, Schless SH, Aertbeliën E, Desloovere K, Bruyninckx H. Efficient image based method using water-filled balloons for improving probe spatial calibration in 3D freehand ultrasonography. Ultrasonics 2019;94:124–130.

Daly L, Bourke G. Interpretation and uses of medical statistics. 5th edition New York: Wiley; 2000.

De Beukelaer N, Bar-On L, Hanssen B, Peeters N, Prinsen S, Ortibus E, Desloovere K, Van Campenhout A. Muscle characteristics in pediatric hereditary spastic paraplegia vs. bilateral spastic cerebral palsy: An exploratory study. Front Neurol 2021;12:1–10.

Fry NR, Gough M, Shortland AP. Three-dimensional realisation of muscle morphology and architecture using ultrasound. Gait Posture 2004;20:177–182.

Fry NR, Gough M, McNee AE, Shortland AP. Changes in the volume and length of the medial gastrocnemius after surgical recession in children with spastic diplegic cerebral palsy. J Pediatr Orthop 2007;27:769–774.

Fukunaga T, Miyatani M, Tachi M, Kouzaki M, Kawakami Y. Muscle volume is a major determinant of joint torque in humans. Acta Physiol Scand 2001;172:249–255.

Haberfehlner H, Jaspers RT, Rutz E, Becher JG, Harlaar J, van der Sluijs JA, Witbreuk MM, Romkes J, Freslier M, Brunner R, Maas H, Buizer AI. Knee moment-angle characteristics and semitendinosus muscle morphology in children with spastic paresis selected for medial hamstring lengthening. PLoS One 2016a;11:e0166401.

Haberfehlner H, Maas H, Harlaar J, Becher JG, Buizer AI, Jaspers RT. Freehand three-dimensional ultrasound to assess semitendinosus muscle morphology. J Anat 2016b;229:591–599.

Haberfehlner H, Jaspers RT, Rutz E, Harlaar J, Van Der Sluijs JA, Witbreuk MM, van Hutten K, Romkes J, Freslier M, Brunner R, Becher JG, Maas H, Buizer AI. Outcome of medial hamstring lengthening in children with spastic paresis: A biomechanical and morphological observational study. PLoS One 2018;13:e0192573.

Handsfield GG, Meyer CH, Abel MF, Blemker SS. Heterogeneity of muscle sizes in the lower limbs of children with cerebral palsy. Muscle Nerve 2016;53:933–945.

Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15:155–163.

Kot BCW, Sin DMH, Ying M. Evaluation of the accuracy and reliability of two 3-dimensional sonography methods in volume measurement of small structures: An in vitro phantom study. J Clin Ultrasound 2009;37:82–88.

Lee M, Ko Y, Myong SooK SHin M, Lee W. The effects of progressive functional training on lower limb muscle architecture and motor function in children with spastic cerebral palsy. J Phys Ther Sci 2015;27:1581–1584.

MacGillivray TJ, Ross E, Simpson HAHRW, Greig CA. 3D freehand ultrasound for in vivo determination of human skeletal muscle volume. Ultrasound Med Biol 2009;35:928–935.

Malaiya R, McNee AE, Fry NR, Eve LC, Gough M, Shortland AP. The morphology of the medial gastrocnemius in typically developing children and children with spastic hemiplegic cerebral palsy. J Electromyogr Kinesiol 2007;17:657–663.

McMahon JJ, Turner A, Comfort P. Within- and between-session reliability of medial gastrocnemius architectural properties. Biol Sport 2016;33:185–188.

McNee AE, Gough M, Morrissey MC, Shortland AP. Increases in muscle volume after plantarflexor strength training in children with spastic cerebral palsy. Dev Med Child Neurol 2009;51:429–435.

Moreau NG, Holthaus K, Marlow N. Differential adaptations of muscle architecture to high-velocity versus traditional strength training in cerebral palsy. Neurorehabil Neural Repair 2013;27:325–334.

Mozaffari MH, Lee WS. Freehand 3-D ultrasound imaging: A systematic review. Ultrasound Med Biol 2017;43:2099–2124.

Noble JJ, Charles-Edwards GD, Keevil SF, Lewis AP, Gough M, Shortland AP. Intramuscular fat in ambulant young adults with bilateral spastic cerebral palsy. BMC Musculoskelet Disord 2014a;15:236.

Noble JJ, Fry NR, Lewis AP, Keevil SF, Gough M, Shortland AP. Lower limb muscle volumes in bilateral spastic cerebral palsy. Brain Dev 2014b;36:294–300.

Noorkoiv M, Theis N, Lavelle G. A comparison of 3D ultrasound to MRI for the measurement and estimation of gastrocnemius muscle volume in adults and young people with and without cerebral palsy. Clin Anat 2019;32:319–327.

Obst SJ, Boyd R, Read F, Barber L. Quantitative 3-D ultrasound of the medial gastrocnemius muscle in children with unilateral spastic cerebral palsy. Ultrasound Med Biol 2017;43:2814–2823.

Palisano R, Rosenbaum P, Walter S, Russell D, Wood E, Galuppi B. Development and reliability of a system to classify gross motor function in children with cerebral palsy. Dev Med Child Neurol 2008;39:214–223.

Pitcher CA, Elliott CM, Fausto A, Valentine JP, Stannage K, Reid SL. Ultrasound characterization of medial gastrocnemius composition in children with spastic cerebral palsy. Muscle Nerve 2015;52:397–403.

Pillen S, Tak RO, Zwarts MJ, Lammens MMY, Verrijp KN, Arts IMP, van der Laak JA, Hoogerbrugge PM, van Engelen BGM, Verrips A. Skeletal Muscle Ultrasound: Correlation Between Fibrous Tissue and Echo Intensity. Ultrasound Med Biol 2009;35:443–446.

Schillebeeckx F, De Groef A, De Beukelaer N, Desloovere K, Verheyden G, Peers K. Muscle and tendon properties of the spastic lower leg after stroke defined by ultrasonography: A systematic review. Eur J Phys Rehabil Med 2020. doi: 10.23736/S1973-9087.20.06462-X. Accessed 16 April 2021. [e-pub ahead of print]

Schless SH, Hanssen B, Cenni F, Bar-On L, Aertbelien E, Molenaers G, Desloovere K. Estimating medial gastrocnemius muscle volume in children with spastic cerebral palsy: A cross-sectional investigation. Dev Med Child Neurol 2018a;60:81–87.

Schless SH, Cenni F, Bar-On L, Hanssen B, Kalkman B, O'brien T, Aertbeliën E, Van Campenhout A, Molenaers G, Desloovere K. Medial gastrocnemius volume and echo-intensity after botulinum neurotoxin A interventions in children with spastic cerebral palsy. Dev Med Child Neurol 2018b;61:783–790.

Schless SH, Cenni F, Bar-On L, Hanssen B, Goudriaan M, Papageorgiou E, Aertbeliën E, Molenaers G, Desloovere K. Combining muscle morphology and neuromotor symptoms to explain abnormal gait at the ankle joint level in cerebral palsy. Gait Posture 2019;68:531–537.

Silvestri E, Muda A, Orlandi D. Ultrasound anatomy of lower limb muscles: A practical guide. Cham: Springer; 2015.

Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. Stat Med 1998;17:101–110.

Watanabe Y, Yamada Y, Fukumoto Y, Ishihara T, Yokoyama K, Yoshida T, Miyake M, Yamagata E, Kimura M. Echo intensity obtained from ultrasonography images reflecting muscle strength in elderly men. Clin Interv Aging 2013;8:993–998.

Weir JP. Quantifying test−retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res 2005;19:231–240.

Weller R, Pfau T, Ferrari M, Griffith R, Bradford T, Wilson A. The determination of muscle volume with a freehand 3D ultrasonography system. Ultrasound Med Biol 2007;33:402–407.

Willerslev-Olsen M, Choe Lund M, Lorentzen J, Barber L, Kofoed-Hansen M, Nielsen JB. Impaired muscle growth precedes development of increased stiffness of the triceps surae musculotendinous unit in children with cerebral palsy. Dev Med Child Neurol 2018;60:672–679.

Williams SA, Elliott C, Valentine J, Gubbay A, Shipman P, Reid S. Combining strength training and botulinum neurotoxin intervention in children with cerebral palsy: The impact on muscle morphology and strength. Disabil Rehabil 2012;35:596–605.

Young HJ, Jenkins NT, Zhao Q, Mccully KK. Measurement of intramuscular fat by muscle echo intensity. Muscle Nerve 2015;52:963–971.