Check for updates

Critical evaluation of commonly used methods to determine the concordance between sonography and magnetic resonance imaging: A comparative study

Konstantin Warneke[1]*, Michael Keiner[2],
Lars Hubertus Lohmann[3], Anna Brinkmann[4], Andreas Hein[4],
Stephan Schiemann[1] and Klaus Wirth[5]

[1]Institute for Exercise, Sport and Health, Leuphana University, Lüneburg, Germany, [2]Department of Exercise Science, German University of Health and Sport, Ismaning, Germany, [3]Institute of Sport Science, Carl von Ossietzky University of Oldenburg, Oldenburg, Germany, [4]Assistance Systems and Medical Device Technology, Carl von Ossietzky University of Oldenburg, Oldenburg, Germany, [5]Department for Sport Science, University of Applied Sciences Wiener Neustadt, Wiener Neustadt, Austria

Introduction: An increasing number of studies investigate the influence of training interventions on muscle thickness (MT) by using ultrasonography. Ultrasonography is stated as a reliable and valid tool to examine muscle morphology. Researches investigating the effects of a training intervention lasting a few weeks need a very precise measurement since increases in MT can be assumed as small. Therefore, the aim of the present work was to investigate the concordance between MT via sonography and muscle cross-sectional area (MCSA) determined via MRI imaging (gold standard) in the calf muscle.

Methods: Reliability of sonography measurement and the concordance correlation coefficient, the mean error (ME), mean absolute error (MAE) and the mean absolute percentage error (MAPE) between sonography and MRI were determined.

Results: Results show intraclass correlation coefficients (ICC) of *0.88−0.95* and MAPE of *4.63−8.57%*. Concordance between MT and MCSA was examined showing $\rho = 0.69−0.75$ for the medial head and *0.39−0.51 c* for the lateral head of the gastrocnemius. A MAPE of 15.88−19.94% between measurements were determined. Based on this, assuming small increases in MT due to training interventions, even with an ICC of *0.95*, MAPE shows a high error between two investigators and therefore limited objectivity.

Discussion: The high MAPE of *15.88−19.94%* as well as CCC of $\rho c = 0.39−0.75$ exhibit that there are significant differences between MRI and sonography. Therefore, data from short term interventions using sonography to detect changes in the MT should be handled with caution.

Key points

The aim of this study was to examine the measurement error between the determination of muscle thickness using sonography and of muscle cross-sectional area using MRI measurement. Results show measurement errors of sonography equal to the expected enhancements in muscle thickness following commonly used training interventions over periods of several weeks. Consequently, assuming increases of 5–10% of muscle cross-sectional area and/or muscle thickness, the use of sonography should be questioned.

Introduction

Increasing muscle mass is of high importance in (elite) sports (Del Vecchio et al., 2019; Kordi et al., 2020; Zaras et al., 2021), prevention – especially in age-related diseases such as sarcopenia (English and Paddon-Jones, 2010; Lopes et al., 2019; Vikberg et al., 2019) – and rehabilitation of orthopedic indications (Wada et al., 2020). Consequently, several training methods aim to improve muscle mass which is usually measured *via* muscle thickness (MT) or muscle cross-sectional area (MCSA) (Schoenfeld et al., 2016; Simpson et al., 2017; Wackerhage et al., 2019; May et al., 2021), also in rehabilitative settings (Guthrie et al., 2012; Larivière et al., 2019; Padulo et al., 2020). Sarto et al. (2021) pointed out promising applications of sonography in (elite) sports settings. Imaging *via* sonography is often used to pinpoint injuries and muscular imbalances (Connell et al., 2004, 2006; Balius et al., 2012). But, especially if the aim is to determine changes from pre- to post-test, e.g., in sarcopenia (Rustani et al., 2019), or following a training intervention (Ticinesi et al., 2017), an exact determination of possible changes in morphological parameters is requested. This is underlined by documented increases of moderate effect sizes in MCSA or MT (5.56% – 8.02%, $d = 0.36$–0.58) (Schoenfeld et al., 2016; Amirthalingam et al., 2017; Coratella et al., 2018; Prestes et al., 2019; Matos et al., 2022) in different muscle groups. In addition, there were even higher increases of up to 17.78% (from 18.0 ± 4.7 mm to 21.1 ± 16 mm, $d = 0.66$) (Evangelista et al., 2019) in untrained individuals following an 8-week whole body resistance training intervention and of 13.06% (from 26.8 ± 5.9 mm to 30.3 ± 5.9 mm) with $d = 0.59$ (Ozaki et al., 2020) in the leg muscles of untrained individuals after a 12-week training intervention – both measured *via* sonography. Using MRI, there were also increases in MSCA of 4 – 6.1% in the quadriceps, (Athiainen et al., 2005; Souza et al., 2014; Watanabe et al., 2014; Tavares et al., 2017) and 7.4% for muscle volume with $d = 0.38$ (Wirth et al., 2007) in the biceps brachii muscle.

Accordingly, a high concordance between measurement procedures with very low measurement error is mandatory to rule out the possibility that measured changes could be explained due to measurement error. To investigate MT, sonography is commonly used (Sarto et al., 2021) and is described as a valid and reliable assessment with inter-day reliability of an intraclass correlation coefficient (ICC) of *0.72–0.99* (Wong et al., 2013; Rosenberg et al., 2014) and intra-day reliability of ICC = *0.97–0.99* in the multifidus lumborum and the gastrocnemius muscle, as well as very high inter-rater- and intra-rater reliability of ICC = *0.78–0.94* (Wallwork et al., 2007; Teyhen and Koppenhaver, 2011; Chiaramonte et al., 2019; Betz et al., 2021) measured in the quadriceps and multifidus muscle. However, reviews by English et al. (2012) and Hebert et al. (2009) only showed moderate reliability for intra- and inter-day reliability *(ICC = 0.62–0.97)* when investigating muscle size *via* sonography. Both reviews pointed out partially low quality, high bias of sonography measurements, and heterogeneity of reliability in included studies. Accordingly, Barotsis et al. (Barotsis et al., 2020) provided low to high ICC when performing sonography measurements three times within 24 hours of *0.30–0.99*, dependent on the muscle examined, with the lowest ICC of *0.3* in the massester muscle and the highest ICCs *0.99* in the arm and the calf muscles (Panidi et al., 2021; Yahata et al., 2021; Warneke et al., 2022). Kim et al. (2011) also showed Pearson correlation coefficients and ICC for determining muscle size *via* sonography of *0.43–0.53* in the supraspinatus. To evaluate and compare correlations between different studies, a high degree of standardization is required. However, the literature reveals challenges in standardizing sonography as results are highly dependent on the localization of the measurement (English et al., 2012) and the pressure applied by the examiner (Sarto et al., 2021). Magnetic resonance imaging (MRI) is stated as gold standard when investigating MCSA (Bemben, 2002) and predicting injuries and recovery (Connell et al., 2004). MRI is often used to evaluate sonography's validity by comparison or correlation analyses (Bemben, 2002).

Some studies show high correlations of up to $r = 0.97$ (Bemben, 2002; Thomaes et al., 2012; Palmer et al., 2015) between both methods, depending on measured muscles (Giles et al., 2015), ranging from $r = 0.37$–0.97. There were correlations of $r = 0.72$–0.97 for muscle of the shoulder (Dupont et al., 2001; Khoury et al., 2008), $r = 0.82$–0.88 for the hip muscles (Mendis et al., 2010), $r = 0.71$–0.94 for the hamstrings (Palmer et al., 2015) and $r = 0.88$–0.94 for the forearm muscles (Abe et al., 2018). However, using traditional correlation calculations (Pearson, Spearman, ICC) to examine the concordance of two measurements must be questioned (Lin, 1989; Grouven et al., 2007). Lin (1989) suggested using the concordance correlation coefficient (CCC) to explore the reproducibility of two methods when evaluating measurement devices. Furthermore, when examining parameters gathered by two measurement techniques, there should be a very low level of variance between those techniques. Bland-Altman (BA) analysis (Giavarina, 2015; Dogan, 2018) is recommended to evaluate the level of variance between two testing methods (Grouven et al.,

2007). Generally, ICC and/or Pearson correlations classified as high (>0.8) are commonly used to justify sonography as an alternative method to MRI (Mendis et al., 2010; Yi et al., 2012; Betz et al., 2021). Abe et al. (2015) and Franchi et al. (2018) assumed that MT reflects the cross-sectional area in the lower legs in the lower extremity, including the quadriceps and triceps surae. Because of comparatively low morphological increases due to training interventions of up to 8 weeks, a highly sensitive and accurate determination of MCSA and MT is required. No statement about the mean error of the measurement is given using correlative calculations, which should be involved to the evaluation of a measurement device to estimate the precision of a measurement. However, since there are conflicting results regarding the reliability of sonography in determining muscle morphology, an evaluation of both methods is required. Consequently, this study critically evaluates the requested aspects (Abe et al., 2015; Ticinesi et al., 2018): First, the reliability and the agreement of sonography measurements between two raters. Second, the detection of the reproducibility and concordance between MT measured *via* sonography and MCSA measured *via* MRI in the calf muscle.

Methods

To examine the concordance between both methods, 96 MRI and sonography values were evaluated. The literature shows differing reliabilities dependent on muscle groups (Giles et al., 2015; Barotsis et al., 2020) with the leg muscles exhibiting high reliability values with ICCs of up to *0.99* in the plantar flexors (Panidi et al., 2021; Yahata et al., 2021; Warneke et al., 2022). Thus, the gastrocnemius muscle was used for this investigation. Furthermore, sonography images were collected from two different investigators to determine the inter-rater reliability of the sonography investigations. Correlations were determined for both investigators for sonography. Furthermore, correlations were calculated between determined MT *via* sonography and MCSA measured *via* the gold standard method MRI.
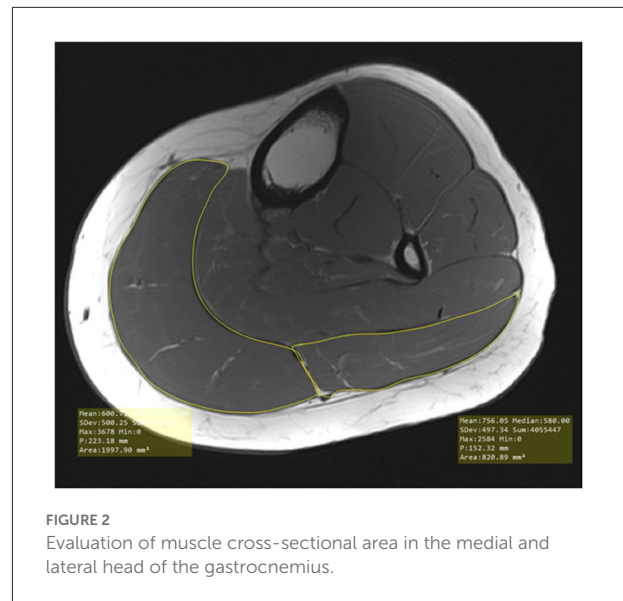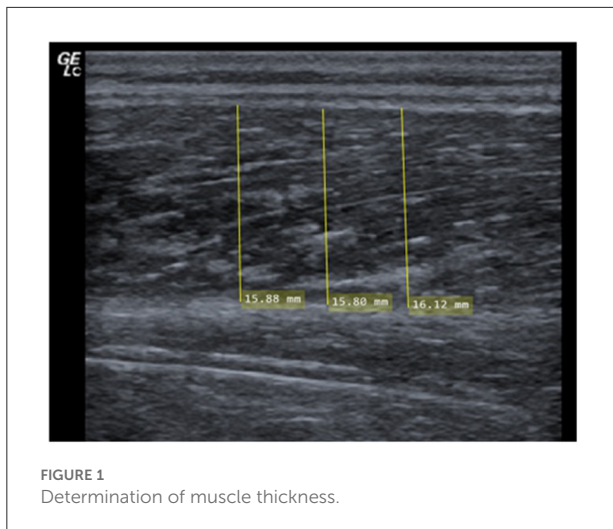
Subjects

Forty eight young healthy subjects (male: 36, female: 12, age: 28.22 ± 5.26 years, height: 181.04 ± 9.58 cm, weight: 81.78 ± 15.51 kg) were recruited from the university campus. Subjects with implants or protheses, claustrophobia or anxiety were excluded from the study. All participants were informed about the experimental risks and provided written informed consent to participate in the present study. Furthermore, approval for this study was obtained from the institutional review board (Carl von Ossietzky University of Oldenburg, No.121-2021). The study

was performed with human participants in accordance with the Helsinki Declaration.

Sonography

MT is defined as the linear, perpendicular distance between the two linear borders of the skeletal muscle and was obtained by averaging three measurements across the proximal, central and distal portions of the obtained ultrasound images (Franchi et al., 2018; Sarto et al., 2021). Two investigators independently evaluated MT using the image processing software MicroDicom (Sofia, Bulgaria). The objectivity of the evaluators was determined as high *(r = 0.87)*. In the literature, high-reliability values of up to *r = 0.9* for determining MT *via* ultrasound for intra-day reliability (Nabavi et al., 2014; Cuellar et al., 2017) and ICC values of up to *0.97* for inter-day reliability are considered as high (König et al., 2014; Rahmani et al., 2019). In the plantar flexors, using sonography to determine MT showed high reliability with ICC of up to *0.99* (Rosenberg et al., 2014; Panidi et al., 2021; Yahata et al., 2021; Warneke et al., 2022). Three images were evaluated for each muscle examined to reduce the standard error. A reduction of the standard error by 50% can be assumed using this procedure (Koppenhaver et al., 2009; Teyhen and Koppenhaver, 2011).

MT was examined in the medial and lateral head of the gastrocnemius. Measurements were performed using a two-dimensional B-mode ultrasound (Mindray Diagnostic Ultrasound System). A linear transducer with a standardized frequency of 12–13 MHz was used to record images of both heads of the gastrocnemius. Each participant was placed in a prone position on a table with the feet hanging down at the end to ensure no contraction in the calf muscles. Subsequently, the sonographer identified the proximal and distal landmark of the lateral gastrocnemius for each participant and measurement (Perkisas et al., 2021). 30% of the distance between the articular cleft of the knee joint to the most lateral top of the lateral malleolus was used to place the transducer (Perkisas et al., 2021). The muscle belly was determined as the center of the muscle between its medial and lateral borders where the maximal MCSA can be assumed (Fukunaga et al., 1992; May et al., 2021). In addition, the image plane is best aligned with the muscle's fascicles including minimal fascicle curvature (Bénard et al., 2009; Raj et al., 2012; May et al., 2021). To improve acoustic coupling and to reduce the transducer's pressure on the skin before starting the measurement, a transmission gel was applied. Next, the investigators ensured that the superficial and deep aponeuroses were as parallel as possible by holding and rotating the transducer around the sagittal-transverse axis to the determined point on the skin without compressing the muscle. Hence, the visibility of the fascicles as continuous striations from one aponeurosis to the other was optimized (see Figure 1).

FIGURE 1
Determination of muscle thickness.

FIGURE 2
Evaluation of muscle cross-sectional area in the medial and lateral head of the gastrocnemius.

MRI measurement

MRI was performed at the Neuroimaging Unit of the Carl von Ossietzky University of Oldenburg using a 3T Siemens Magnetom Prisma MRI with a T1-weighted turbo-spin-echo sequence (40 slices, slice thickness = 7 mm, TR = 1600 ms, TE = 14 ms, voxel size = 0.4 x 0.4 mm², FOV = 150 x 150, distance factor = 20%, flip angle = 150°, TA = 8:16 min) with a combination of the body and spine coil. Each participant was placed on the back and the measurement was performed first on the left leg, immediately followed by the right leg. The evaluation of MRI images and therefore examination of MCSA was performed by bordering the fascia layers of the lateral and the medial head of the gastrocnemius (see Figure 2) with MicroDicom (Sofia, Bulgaria) by two investigators independently from each other and anonymized for participant and group. Examination of images started from the first image distal of the knee joint where a clear bordering of the muscle could be seen to the transition from the muscle to the tendon. For evaluation of MCSA the mean of the three highest MCSA values in the lateral and the medial head of the gastrocnemius were used to minimize potential error of location (Koppenhaver et al., 2009; Teyhen and Koppenhaver, 2011). Reliability of MRI measurements can be assumed as very high with $r = 0.99$ (Wirth et al., 2007; Wang et al., 2021).

Data analysis

The data was analyzed using SPSS 28.0 (IBM, Ehningen, DE, Germany) and graphics were produced with "R." The significance level for all statistical tests was set at $p < 0.05$. The descriptive statistics for all measures are presented as the mean (M) ± standard deviation (SD). Values were obtained from an intervention study consisting of pre-test and post-test values. To determine significant differences in the correlation coefficients between subgroups (different measurement times), the data were z-transformed according to the Fisher method (Ferreira and Zwinderman, 2006). Reliability was determined for intra-day and inter-day reliability. Calculation of ICC as well as coefficients of variability (CV) between the best and the second-best value of MT within 1 day (intra-day reliability), the best values between two consecutive days (inter-day reliability) and the best value of investigator one and investigator two (investigator objectivity) were calculated. Two investigators performed this statistical procedure to determine objectivity in the calculation. Furthermore, a two-tailed Pearson correlation was determined between the best value of measured MT and the maximal MCSA determined by MRI. Then, reproducibility and concordance were determined between sonography measurements conducted by both investigators by calculating Lin's CCC. Lin (1989) suggested using CCC for the evaluation of medical devices if the aim is to examine reproducibility which is, in fact, not given by using Pearson correlations (Lin, 1989; Koch and Spörl, 2007; Kwiecien et al., 2011). Furthermore, the mean error (ME) between both testing methods was calculated by

$$ME = \frac{1}{n} * \sum_{i=1}^{n} (x_i - y_i),$$

the mean absolute error (MAE) by

$$MAE = \frac{1}{n} * \sum_{i=1}^{n} |x_i - y_i|,$$

TABLE 1 Intra-day reliability with intra class correlations and coefficients of variability.

Parameter	ICC (95% CI)	CV (in %)
Gastrocnemius lateralis (investigator 1)	0.88 (0.8–0.93)	4.19 ± 5.3
Gastrocnemius medialis (investigator 1)	0.95 (0.90–0.97)	2.86 ± 2.37
Gastrocnemius lateralis (investigator 2)	0.90 (0.84–0.95)	4.42 ± 5.01
Gastrocnemius medialis (investigator 2)	0.937 (0.89–0.97)	3.15 ± 2.45

ICC, intraclass correlation coefficient; CV, coefficient of variability.

TABLE 2 Descriptive statistics of MT measured with sonography and MCSA measured with MRI.

Parameter	M ± SD	Minimum	Maximum
MT G.Medialis (1)	18.75 ± 3.02	12.68	26.14
MT G.Lateralis (1)	14.60 ± 2.66	9.41	23.21
MT G.Medialis (2)	18.89 ± 3.17	12.74	27.61
MT G.Lateralis (2)	14.93 ± 2.78	10.20	22.63
MCSA G. Medialis	1,713.20 ± 468.24	741.33	3,203.00
MCSA G. Lateralis	1,020.99 ± 245.75	503.65	1,564.00

MT, muscle thickness; MCSA, muscle cross-sectional area; G. Medialis, medial head of the gastrocnemius; G. Lateralis, lateral head of the gastrocnemius; (1), first investigator; (2), second investigator.

and the mean absolute percentage error (MAPE) by

$$MAPE = \frac{100\%}{n} * \sum_{i=1}^{n} \left| \frac{x_i - y_i}{x_i} \right|.$$

To determine the reproducibility of MRI, CCC was calculated to evaluate whether or to what extent MT and MCSA measure the same parameter. Additionally, MAE and MAPE are provided. For this reason and because different units in sonography (MT in mm) and MRI (MCSA in mm²) were used, further calculation was done using z-transformed data:

$$x_i \rightarrow \frac{x_i - \bar{x}_n}{\sigma}.$$

Results

Intra-day reliability for both investigators

Intra-day reliability of sonography measurement of both investigators is provided in Table 1.

Concordance and inter-rater reliability in sonography imaging between both investigators

For the medial head of the gastrocnemius inter-rater reliability between investigator 1 and investigator 2 can be assumed as high with [$r = 0.93$ (0.90–0.96, 95% CI), ICC = 0.93 (0.90–0.95, 95% CI), CV = 3.26 ± 2.68%, and $\rho_c = 0.93$ (0.9–0.95, 95% CI)].

For the lateral head of the gastrocnemius inter-rater reliability between investigator 1 and investigator 2 can be assumed as high with $r = 0.834$ [0.758–0.887, 95% CI], ICC = 0.833 [0.76–0.89, 95% CI] CV of 5.92 ± 5.43% and $\rho_c = 0.8$ [0.67–0.88, 95% CI].

Investigating the ME, MAE and MAPE for the medial head of the gastrocnemius results showed a ME = −0.14 mm, a MAE = 0.88 mm and a MAPE = 4.63%, while in the lateral head of the gastrocnemius, there was a ME = −0.33 mm, MAE = 1.20 mm, and MAPE = 8.57%.

Figure 3 shows the CCC for sonography measurement between investigator 1 und investigator 2 in the lateral head of the gastrocnemius (a) and medial head of the gastrocnemius (b).

Concordance between sonography and MRI

Descriptive statistics of MT and MCSA is provided in Table 2. Using z-transformed values, there is a MAE of 0.70 and ME of −0.15 between MRI and sonography in the gastrocnemius medialis using values of the first investigator and MAE of 0.79 and ME of −0.15 when using values of the second investigator.

Using z-transformed values between MRI and sonography, there is a MAE in the gastrocnemius lateralis of 0.98 and ME of −0.18 using values from first investigator and MAE of 0.99 and ME of −0.18 when using values of the second investigator.

To determine the relationship between the real values of the sonography and MRI measurement, the linear trend line was taken for MRT measure (y) as a function of sonography measure (x):

Gastrocnemius medialis:
1. investigator: $y = f(x) = 116.738 * x − 476.07$
2. Investigator: $y = f(x) = 102.245 * x − 218.196$
Gastrocnemius lateralis:
1. investigator: $y = f(x) = 46.844 * x + 336.942$
2. Investigator: $y = f(x) = 34.735 * x + 502.438$

The differences for all x between f (x) and the associated MRT-values were determined and absolute and mean values were used for further calculation. Based on this, when using values from investigator 1 for MT in the medial head of the gastrocnemius there was a *ME = 0.0001, MAE = 251.483* with a *maximum = 754.988* and a *MAPE = 15.88%* with a *maximum = 72%* between MCSA and MT. Using values from investigator 2, there was a *ME = 0.0036*, a *MAE = 268.665* with a *maximum = 1017.608* and a *MAPE = 17.066%* with a *maximum = 77.52%*.

Determining the ME, MAE, and MAPE between MCSA and MT in the lateral head of the gastrocnemius showed a *ME =*

FIGURE 3
Concordance correlation coefficient between investigator 1 and investigator 2 for determining muscle thickness in the medial head **(A)** and the lateral head **(B)** of the gastrocnemius.
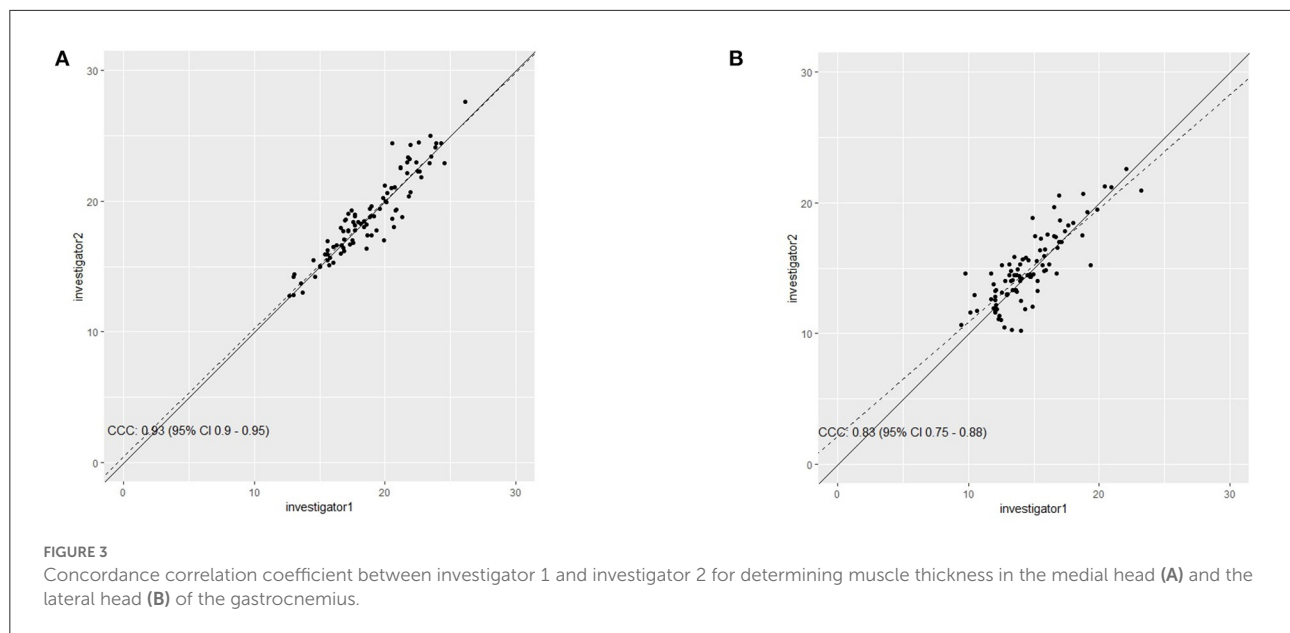
0.0003, a $MAE = 177.201$ with a $maximum = 434.459$ and a $MAPE = 19.14\%$ with a $maximum = 77.57\%$ using MT values from investigator 1. Using MT values from investigator 2 there was a $ME = 0.0005$, a $MAE = 184.207$ with a $maximum = 491.070$ and a $MAPE = 19.94\%$ with a $maximum = 94.76\%$.

All MRT-values are reported in mm² except for MAPE.

Concordance between both measurements was calculated with CCC by Lin (1989) and is plotted for the lateral head of the gastrocnemius in Figure 4 for the first investigator (a) and the second investigator (b) and for the medial head of the gastrocnemius in Figure 5.
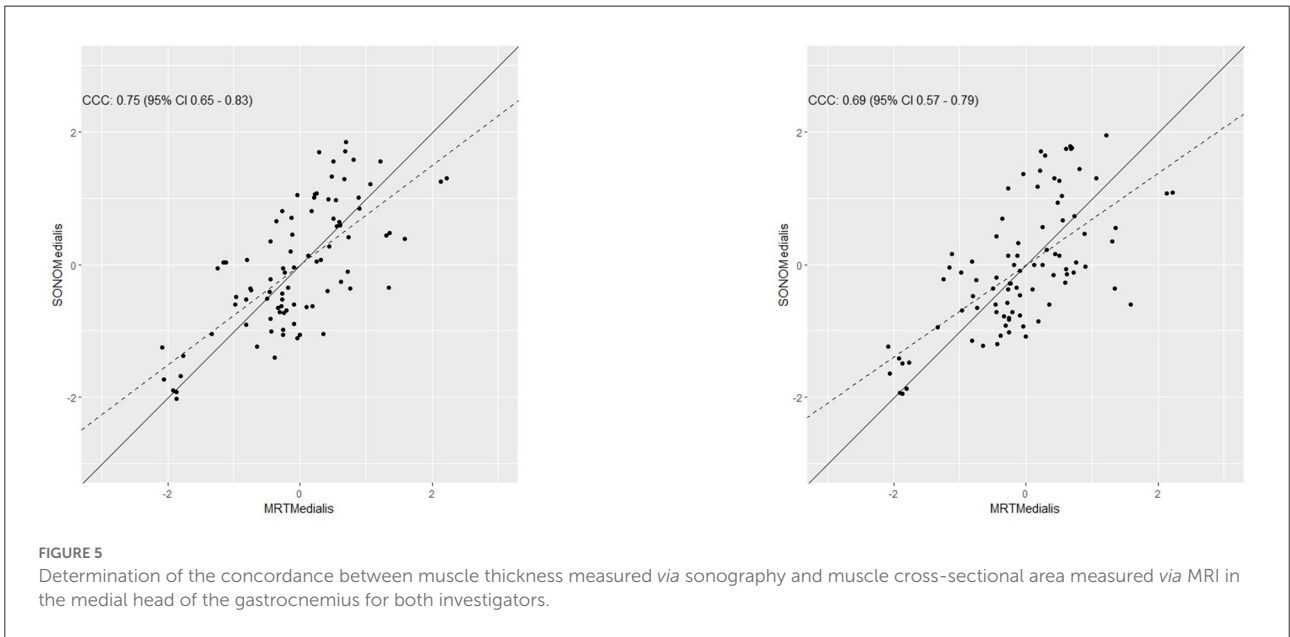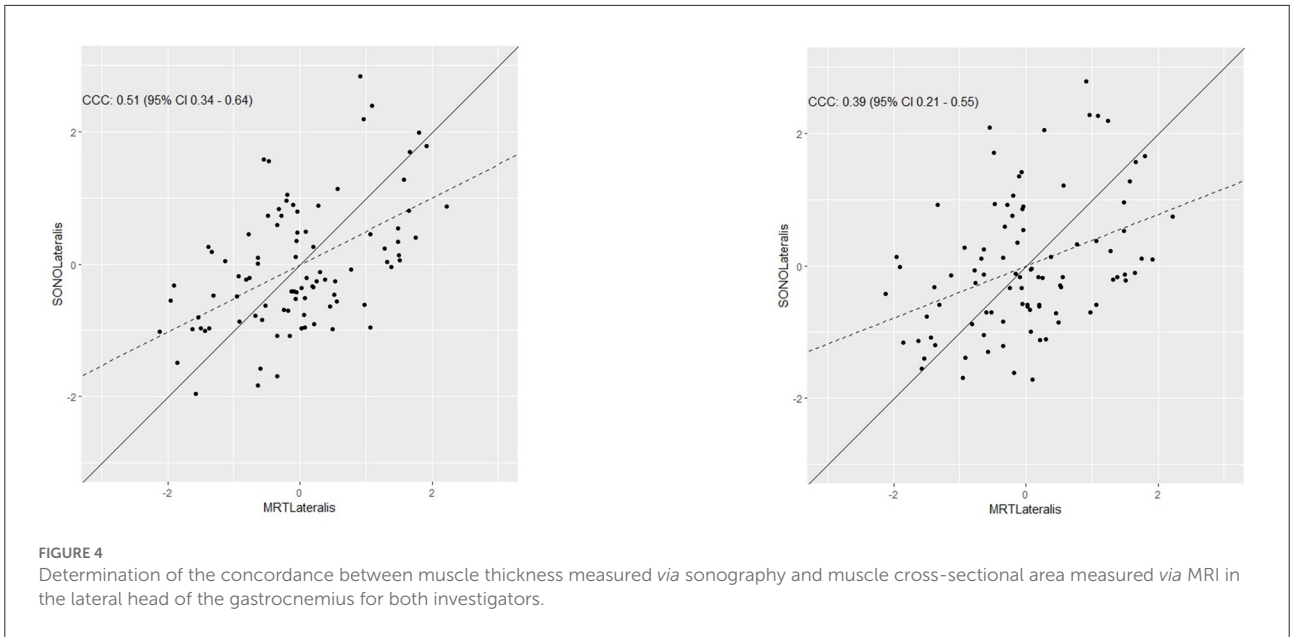
Pearson correlations coefficients (r) and CCC (ρ_c) were calculated and are provided in Table 3.

Discussion

The aim of present work was to investigate the concordance between MT *via* sonography and MCSA determined *via* MRI (gold standard) in the calf muscle. In general, ICC and/or Pearson correlations are used to argue for sonography as an alternative method to MRI (Mendis et al., 2010; Yi et al., 2012; Betz et al., 2021). MT is assumed to reflect the cross-sectional area in the lower legs (Abe et al., 2015; Franchi et al., 2018). However, since there are conflicting results regarding the reliability of sonography in determining muscle morphology, an evaluation of both methods is required. Consequently, this study aimed to critically evaluate the following two aspects: First, the reliability and concordance of sonography measurements between two investigators. Second, the detection of the reproducibility and concordance between MT measured *via* sonography and MCSA measured *via* MRI in the calf muscle.

Critical evaluation of sonography measurement to examine muscle thickness

In literature, there are studies showing correlation coefficients of $r = 0.37–0.97$ between sonography and MRI (Bemben, 2002; Thomaes et al., 2012; Palmer et al., 2015), arguing that sonography is a reliable and valid alternative to determine morphological changes following training interventions or muscular disuse. Based on Pearson correlation coefficients, some authors suggest using sonography to determine hypertrophy or atrophy following training interventions or sarcopenia (Rustani et al., 2019). The present study found high correlations of $r = 0.83–0.93$ and ICC values between two investigators with high ICC for intra-day reliability $ICC = 0.83–0.93$ of sonography as well. However, English et al. (2012) stressed some methodological issues of the included studies in their review, pointing out an overestimation of reliability in sonography. First, problems of the listed studies arise from inadequate statistical analyses. Most studies state that ICC and correlation coefficients calculate the reliability of sonography measurements. However, the classification into "high," "moderate," and "low" as well as the following interpretation seem inaccurate and should therefore be questioned. Studies point out reliability as "good overall" with inter- and intra-day reliability of $ICC = 0.67–0.99$ (Bentman et al., 2010; Wong et al., 2013; Rosenberg et al., 2014) and very high inter- and intra-rater reliability with $ICC = 0.77–0.94$ (Wallwork et al., 2007; Teyhen and Koppenhaver, 2011; König et al., 2014; Temes et al., 2014; Chiaramonte et al., 2019; Betz et al., 2021). Another systematic review from 2017 found intra- and inter-rater reliabilities

FIGURE 4
Determination of the concordance between muscle thickness measured *via* sonography and muscle cross-sectional area measured *via* MRI in the lateral head of the gastrocnemius for both investigators.

FIGURE 5
Determination of the concordance between muscle thickness measured *via* sonography and muscle cross-sectional area measured *via* MRI in the medial head of the gastrocnemius for both investigators.

TABLE 3 Comparison between Pearson correlation coefficient and the concordance analysis for lateral and medial head of the gastrocnemius for both investigators.

	MT G.Lateralis MCSA G.Lateralis	MT G.Medialis MCSA G.Medialis
Investigator 1	$r = 0.57$ (0.34–0.65)	$r = 0.75$ (0.65–0.83)
	$\rho_c = 0.51$ (0.34–0.64)	$\rho_c = 0.75$ (0.65–0.83)
Investigator 2	$r = 0.39$ (0.20–0.55)	$r = 0.69$ (0.57–0.79)
	$\rho_c = 0.39$ (0.21–0.55)	$\rho_c = 0.69$ (0.57–0.79)

MT, muscle thickness; MCSA, muscle cross-sectional area; G. Medialis, medial head of the gastrocnemius; G. Lateralis, lateral head of the gastrocnemius; (1), first investigator; (2), second investigator.

of ICC = *0.45–0.99* examining the morphology of tendons (McAuliffe et al., 2017). Considering a MAPE of 4.6–8.6% with a corresponding inter-rater reliability of ICC = *0.892–0.931*, which are higher than many of the ICCs considered high in the literature, the perception of "high reliability" is strongly biased. Second, problems in standardization within the included studies further limit the confidence of data interpretation (English et al., 2012). It was pointed out that most studies did not provide information on the location and usage of the transducer. Especially by determining the effects of an intervention these information would be mandatory (English et al., 2012). Standardized protocols should be adopted to ensure the quality and comparability of studies (Connolly et al., 2015). Accordingly, using sonography to determine MT and muscle architecture, especially in pre-post comparisons, should be critically questioned (Bentman et al., 2010) because of mentioned limitations in standardization and great subjective influences in the procedure listed by many authors, which can be attributed to e.g., the influence of the pressure applied to the transducer as well as the angle of the transducer (Hebert et al., 2009; Bentman et al., 2010; Connolly et al., 2015). Consequently, the accuracy of reliability calculations can also be questioned since the measurement methodology seems to lack objectivity or missing details on the measurement's repeatability. These problems lead to limited comparability of MT which can be seen in MAPE and MAE between two investigators performing sonography measurement in one participant in a cross-sectional study design. This is of crucial importance when sonography is used to determine changes in MT in pre-post comparisons (Schoenfeld et al., 2017, 2019; Ehsani et al., 2019) considering this study found a MAPE *4.6–8.6%* between investigators for the use of sonography compared to MT percentage increases of *4.5–8%* in listed studies (Athiainen et al., 2005; Watanabe et al., 2014; Schoenfeld et al., 2016; Coratella et al., 2018; Matos et al., 2022). Based on this, these values do not seem to be sufficient to describe sonography as a very precise and adequate measurement device which is requested to determine muscle morphology (Ticinesi et al., 2017). Furthermore, the classification of reliability values and correlation coefficients should be classified according to the purpose of use (Cohen, 1988) and should consider the expected effects of the performed intervention.

Critical evaluation of interpretations from current literature

If sonography imaging is considered "*in the light of methodological limitations [. . .] which may have led to overestimation of reliability indices*" (English et al., 2012, p. 942), a calculation of concordance between sonography and MRI (which is assumed as the gold standard method because of high objectivity in imaging) seems to be questionable in

general, especially using correlation coefficients and ICC. However, if a study aims to determine the concordance between sonography and MRI, correlation coefficients and ICC values are usually calculated (Bemben, 2002; Betz et al., 2021), but considering that correlations only point out the association between variables (Schober and Schwarte, 2018), which only validates to investigate "*the change in the magnitude of 1 variable [. . .] associated with a change in the magnitude of another variable, either in the same or in the opposite direction*" (Schober and Schwarte, 2018, p. 1,763). In the supraspinatus muscle there are correlations coefficients between MT and MCSA of $r = 0.72–0.76$ (Yi et al., 2012). Betz et al. (2021) point out that "*[a] strong predictive positive correlation for ultrasound and magnetic resonance imaging-based measurements of the cross-sectional area was found ($R^2 = 0.793$, $p < 0.001$)*" while Mendis et al. (2010) report ICC values of *0.81–0.89* in different leg muscles to determine the concordance between sonography imaging and MRI. Also, "excellent agreement" between both measures to examine MCSA with $r = 0.96$ and ICC of *0.9–0.96* in the hip muscles (Mayes et al., 2015) and $r = 0.87$ between MT with sonography and MCSA with MRI (Worsley et al., 2014). Giles et al. (2015) point out high correlations in MCSA of $r = 0.73–0.88$ in some parts of the quadriceps for the vastus medialis and but also low correlations with $r = 0.2$ and $r = 0.31$ for other parts as the vastus intermedius. There are also correlations of *0.96* and *0.97* between MRI and sonography in the shoulder (Dupont et al., 2001). The listed studies showed additional limitations. The sample size to determine correlations was small – *eleven* (Mayes et al., 2015) and six participants (Dupont et al., 2001), respectively. In the present study, correlations between MT and MCSA in the plantar flexors were determined showing Pearson correlation coefficients between $r = 0.41–0.72$. However, considering that sonography is an inexpensive and time-economic procedure and could be used as an alternative to MRI, determining Pearson correlation must be stated as invalid because it ignores many parameters especially the level of variance and the expected error between the two methods. Since Pearson correlation coefficients point out a dependency/relationship between two measurements, Lin (1989) suggests using the CCC as it was used in the present study to assess concordance between the measurements. With $\rho_c = 0.39–0.51$ in the lateral head of the gastrocnemius and $\rho_c = 0.69–0.75$ in the medial head of the gastrocnemius the CCC values are lower or equal to the Pearson correlation coefficients with $r = 0.39–0.57$ and $r = 0.69–0.75$. Moreover, ME, MAE, and MAPE should be calculated in this context. When z-transformed data is used, calculation of MAPE seems not to be useful since values on the x-axis close to zero with a corresponding value that is many times higher would lead to a percentage error of over 100% which does not reflect reality. Using ME, negative and positive values counterbalance each other to a large extent, so ME is also not of high value in our setting.

Consequently, MAE values should be recognized as most important. It exhibits values of *177.2–184.21 mm²* in the lateral head and *251.48–268.67 mm²* in the medial head, corresponding to a MAPE of *19.1–19.9%* and *15.9–17.1%*, respectively. Based on mean values of *1,713 mm²* in the medial head in MCSA measurement, MAE and MAPE values of *251.48 mm²* must be recognized as high when a replacement of MRI with sonography is considered. However, on the one hand, high values of MAE in MT measurements do not surprise as this method measures the distance between two points which represents a one-dimensional measurement without any statement about the anatomical shape of a muscle. On the other hand, in MRI the shape of a muscle is used to calculate the area which therefore is a multidimensional evaluation of muscle morphology.

Novelty of the study and comparison to commonly used statistics

In current literature, reliability of sonography was typically determined *via* Pearson correlations and ICC with values of > 0.8 being used to justify the use of sonography as a valid and reliable method to investigate effects of training programs aiming to increase MT and MCSA (Schoenfeld et al., 2016; Simpson et al., 2017; Panidi et al., 2021; Sarto et al., 2021; Yahata et al., 2021; Matos et al., 2022; Warneke et al., 2022). When replacing one measurement procedure with another, it should be assumed that both measure the same parameter, however, there are some significant limitations replacing MRI with sonography, which are of statistical-methodological and content nature. First, correlation coefficients describe a monotonic relationship between two variables "*in which either (1) as the value of 1 variable increases, so does the value of the other variable; or (2) as the value of 1 variable increases, the other variable value decreases*"(Schober and Schwarte, 2018, p. 1,763) while the CCC "*plot the first measurement against the second measurement […], we would like to see, within a tolerable error, that the measurements fall on a 45° line through the origin […]. The Pearson correlation coefficient measures a linear relationship but fails to detect any departure from the 45' line*" (Lin, 1989, p. 255). The Pearson correlation, therefore, does not provide any information about the concordance between two procedures but about their relationship, which is, in fact, not the aim when investigating the possible replaceability of one method with another. Based on this, using correlation coefficients to justify the replaceability of MRI with sonography should be stated as a misinterpretation of statistics and should therefore be avoided. Another well-known method to show the variance between two measurement procedures in the context of medicine is the Bland-Altman Analysis. As the classification of the results also depends on the context, there was, however, no additional benefit of adding the Bland-Altman analysis to the present study.

Additionally, since English et al. (2012) and Hebert et al. (2009) pointed out limitations of studies investigating reliability and stated limited objectivity of sonography, thus, a reevaluation of the usage of sonography seems requested. In accordance with Cohen (Cohen, 1988), the classification in high, moderate and low should be reviewed considering the context.

To the best knowledge, there are few studies investigating the concordance between MRI and sonography using BA analysis and/or the CCC, however, these investigate the quadriceps femoris (Ahtiainen et al., 2010; Ruple et al., 2022). Only Scott et al. (2017) provided first data investigating the concordance of MRI and sonography using BA and CCC for the calf muscle, showing – according to the results of the presented study – "poor" concordance with $\rho_c = 0.37$, while a higher CCC was determined for imaging in the quadriceps. However, listed studies do not provide CCC for the investigation of MT between two investigators and do not assign a ME, MAE and MAPE to the concordance analysis. Scott et al. (2017) stated "Concordance between ultrasound and MRI was excellent in the quadriceps (CCC: 0.78; $P < 0.0001$)", however, considering results presented in this study a CCC of *0.8–0.93* showed corresponding MAPE values of up to *8.5%*.

Consequently, results found in this study show firstly that even ICC and Pearson correlations > 0.9 cannot be deemed "high" in the context of highly sensitive measurement procedures such as sonography because the MAPE and MAE in combination with calculation of the CCC show intolerably high measurement errors. Based on this, sonography to assess muscle hypertrophy in the calf muscle should be handled with care. Secondly, with stated MAPE of up to *20%* and a $\rho_c = 0.39–0.75$ between MCSA investigated *via* MRI and MT investigated *via* sonography, the hypothesized predictability of MCSA using MT (Abe et al., 2015; Franchi et al., 2018) seems questionable. Thirdly, classification of concordance, based on calculated ICC, CCC or BA-analyses should also include the MAE and MAPE, especially when examining the possibility of a replacement of one measurement procedure with another. It can be suggested to combine the above stated parameters when the changes following an intervention are expected to be small. When expecting increases in MT of *5.56–17.78%*, a MAPE of up to *8.5%* between two investigators should be considered too high, even though there were high ICC values.

Limitations

Barotsis et al. (2020) and Giles et al. (2015) showed that reliability of sonography may differ depending on the muscle groups. Since in literature the ICCs and Pearson correlation coefficients were found to be as high as *0.99* in the plantar flexors and no higher coefficients could be detected, this study used imaging procedures in the plantar flexors. However, based on this, there might be limited transferability to other muscle groups, which should be investigated in further research, as the

ICC and Pearson correlation does not provide any information about the MAE and MAPE. Furthermore, sonography and MRI were performed in young and healthy participants. Especially when using sonography, on the one hand, a detrimental influence on the imaging quality can be assumed in participants with high body fat, while, on the other hand, competitive athletes from most sport backgrounds may have lower body fat than "normal" participants which might influence the evaluation (Teyhen and Koppenhaver, 2011; Betz et al., 2021). Based on this, assuming imaging quality might influence the calculated parameters, the recruited participants as well as the resolution of sonography imaging might be of importance for further results. The influence of using higher and lower resolution sonography assessment as well as the inclusion of participants from different performance level as well as different muscles should be investigated in further research, as it can be assumed that differences in body fat and fatty tissue might influence the quality of images and therefore the calculation of error.

Conclusion

Although correlation coefficients as well as ICCs are comparable with previous investigations about reliability and validity of sonography, the presented results show a MAPE between *4.4* and *8.9%* which corresponds to the estimated increases in MCSA and MT. Thus, when measuring hypertrophy following training interventions, a very careful interpretation of data must be requested and potential sources of error in sonography must be kept in mind. Furthermore, the results clearly indicate that even correlation coefficients with $r > 0.9$ cannot be seen as valid indicator of concordance between two testing procedures since correlation coefficients do not examine this issue appropriately. Kwiecien et al. (2011) point out that disrespecting the concordance analysis leads to wrong results. Based on this, using correlation coefficients to examine the concordance between two measurements can be seen as a misinterpretation of results. If the aim of an investigation is the determination of concordance between measurements, a BA-Plot as well as a CCC calculation should at least be added to the analysis (Koch and Spörl, 2007) and interpreted in the light of the respective context, as the classification in high concordance, moderate concordance and low concordance might depend on the system evaluated and the expected adaptations induced by the intervention. Especially when low to moderate effect sizes are assumed (e.g., in elite sports), determining morphological effects should be performed with MRI, as this is deemed the gold standard with minimal limitations regarding objectivity and reliability. Furthermore, literature points out crucial limitations and primarily poor quality of studies examining the reliability of determining MT *via* sonography because information about standardization to reproduce the study design is rarely included (Hebert et al., 2009; English et al., 2012). Additionally, the predictability of MCSA by determining MT *via* sonography

seems to not be sufficiently given, consequently replacing MRI with sonography might be cost- and time-efficient but not feasible. Because of comparatively high differences and errors between measurements in sonography, MRI must be still recognized as the gold standard in determining muscle morphology. Results of sonography imaging to determine morphological changes in longitudinal studies with intervention periods of a few weeks only should be considered very carefully.

Data availability statement

The datasets generated and analyzed during the current study are available from the corresponding author upon request.

Ethics statement

The studies involving human participants were reviewed and approved by Medical Ethics Committee, University of Oldenburg. The patients/participants provided their written informed consent to participate in this study.

Author contributions

KoW, AB, and LL carried out the experiment. KoW and MK performed the analytic calculations. KoW took the lead in writing the manuscript with support from MK, LL, and AB. AB and AH supported the sonography and MRI evaluation. SS supported in result discussion and writing the final version of the manuscript. KlW supervised the project and provided critical feedback to the design of the study and the statistical analysis. All authors contributed to the article and approved the submitted version.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abe, T., Loenneke, J. P., and Thiebaud, R. S. (2015). Morphological and functional relationships with ultrasound measured muscle thickness of the lower extremity: a brief review. *Ultrasound* 23, 166–173. doi: 10.1177/1742271X15587599

Abe, T., Nakatani, M., and Loenneke, J. P. (2018). Relationship between ultrasound muscle thickness and MRI-measured muscle cross-sectional area in the forearm: a pilot study. *Clin. Physiol. Funct. Imaging* 38, 652–655. doi: 10.1111/cpf.12462

Ahtiainen, J. P., Hoffren, M., Hulmi, J. J., Pietikainen, M., Mero, A. A., Avela, J., et al. (2010). Panoramic ultrasonography is a valid method to measure changes in skeletal muscle cross-sectional area. *Eur. J. Appl. Physiol*. 108, 273–279. doi: 10.1007/s00421-009-1211-6

Amirthalingam, T., Mavros, Y., Wilson, G. C., Clarke, J. L., Mitchell, L., Hackett, D. A., et al. (2017). Effects of a modified german volume training programm on muscular hypertroph and strength. *J. Strength Cond. Res*. 31, 3109–3119. doi: 10.1519/JSC.0000000000001747

Athiainen, J. P., Pakarinen, A., Alen, M., Kreamer, W. J., and Häkkinen, K. (2005). Short vs. long rest period between the sets in hypertrophy resistance training: influence on muscle strength, size, and hormonal adaptations in trained men. *J. Strength Cond. Res*. 19, 572–582. doi: 10.1519/00124278-200508000-00015

Balius, R., Pedret, C., Galilea, P., Idoate, F., and Ruiz-Cotorro, A. (2012). Ultrasound assessment of asymmetric hypertrophy of the rectus abdominis muscle and prevalence of associated injury in professional tennis players. *Skeletal. Radiol*. 41, 1575–1581. doi: 10.1007/s00256-012-1429-y

Barotsis, N., Tsiganos, P., Kokkalis, Z., Panayiotakis, G., and Panagiotopoulos, E. (2020). Reliability of muscle thickness measurements in ultrasonography. *Int. J. Rehabil. Res*. 43, 123–128. doi: 10.1097/MRR.0000000000000390

Bemben, M. G. (2002). Use of diagnostic ultrasound for assessing muscle size. *J. Strength Cond. Res*. 16, 103–108. doi: 10.1519/00124278-200202000-00016

Bénard, M. R., Becher, J. G., Harlaar, J., Huijing, P. A., and Jaspers, R. T. (2009). Anatomical information is needed in ultrasound imaging of muscle to avoid potentially substantial errors in measurement of muscle geometry. *Muscle Nerve* 39, 652–665. doi: 10.1002/mus.21287

Bentman, S., O'Sullivan, C., and Stokes, M. (2010). Thickness of the middle trapezius muscle measured by rehabilitative ultrasound imaging: description of the technique and reliability study. *Clin. Physiol. Funct. Imaging* 30, 426–431. doi: 10.1111/j.1475-097X.2010.00960.x

Betz, T. M., Wehrstein, M., Preisner, F., Bendszus, M., and Friedmann-Bette, B. (2021). Reliability and validity of a standardized ultrasound examination protocol to quantify vastus lateralis muscle. *J. Rehabil. Med*. 53, jrm00212. doi: 10.2340/16501977-2854

Chiaramonte, R., Bonfiglio, M., Castorina, E. G., and Antoci, S. A. M. (2019). The primacy of ultrasound in the assessment of muscle architecture: precision, accuracy, reliability of ultrasonography. Physiatrist, radiologist, general internist, and family practitioner's experiences. *Rev. Assoc. Med. Bras*. 65, 165–170. doi: 10.1590/1806-9282.65.2.165

Cohen, J. (1988). *Statistical Power Analysis for Behavioral Sciences, 2nd Edn*. New York, NY: Psychology Press; Taylor and Francis Group.

Connell, D., Ali, K., Javid, M., Bell, P., Batt, M., Kemp, S., et al. (2006). Sonography and MRI of rectus abdominis muscle strain in elite tennis players. *Am. J. Roentgenol*. 187, 1457–1461. doi: 10.2214/AJR.04.1929

Connell, D. A., Schneider-Kolsky, M. E., Hoving, J. L., Malara, F., Buchbinder, R., Koulouris, G., et al. (2004). Longitudinal study comparing sonographic and MRI assessments of acute and healing hamstring injuries. *Am. J. Roentgenol*. 183, 975–984. doi: 10.2214/ajr.183.4.18 30975

Connolly, B., Macbean, V., Crowley, C., Lunt, A., Moxham, J., Rafferty, G. F., et al. (2015). Ultrasound for the assessment of peripheral skeletal muscle architecture in critical illness: a systematic review. *Crit. Care Med*. 43, 897–905. doi: 10.1097/CCM.0000000000000821

Coratella, G., Beato, M., Milanese, C., Longo, S., Limonta, E., Rampichini, S., et al. (2018). Specific adaptations in performance and muscle architecture after weighted jumpsquat vs. body mass squat jump training in recreational soccer players. *J. Strength Cond. Res*. 32, 921–929. doi: 10.1519/JSC.0000000000002463

Cuellar, W. A., Blizzard, L., Callisaya, M. L., Hides, J. A., Jones, G., Ding, C., et al. (2017). Test-retest reliability of measurements of abdominal and multifidus muscles using ultrasound imaging in adults aged 50–79 years. *Musculoskelet. Sci. Pract*. 28.79–84. doi: 10.1016/j.msksp.2016.11.013

Del Vecchio, A., Casolo, A., Negro, F., Scorcelletti, M., Bazzucchi, I., Enoka, R., et al. (2019). The increase in muscle force after 4 weeks of strength training is mediated by adaptations in motor unit recruitment and rate coding. *J. Physiol*. 597, 1873–1887. doi: 10.1113/JP277250

Dogan, N. Ö. (2018). Bland-Altman analysis: a paradigm to understand correlation and agreement. *Turk. J. Emerg. Med*. 18, 139–141. doi: 10.1016/j.tjem.09001

Dupont, A. C., Sauerbrei, E. E., Fenton, P. V., Shragge, P. C., Loeb, G. E., Richmond, F. J., et al. (2001). Real-time sonography to estimate muscle thickness: comparison with MRI and CT. *J. Clin. Ultrasound*. 29, 230–236. doi: 10.1002/jcu.1025

Ehsani, F., Hedayati, R., Bagheri, R., and Jaberzadeh, S. (2019). The effects of stabilization exercise on the thickness of lateral abdominal muscles during standing tasks in women with chronic low back pain: a randomized triple-blinded clinical trial study. *J. Sport Rehabil*. 29, 942–951. doi: 10.1123/jsr.2019-0058

English, C., Fisher, L., and Thoirs, K. (2012). Reliability of real-time ultrasound for measuring skeletal muscle size in human limbs *in vivo*: a systematic review. *Clin. Rehabil*. 26, 934–944. doi: 10.1177/0269215511434994

English, K. L., and Paddon-Jones, D. (2010). Protecting muscle mass and function in older adults during bed rest. *Curr. Opin. Clin. Nutr. Metab. Care*. 13, 34–39. doi: 10.1097/MCO.0b013e328333aa66

Evangelista, A. L., De Souza, E. O., Moreira, D. C. B., Alonso, A. C., Teixeira, C. V. S., Wadhi, T., et al. (2019). Interset Stretching vs. traditional strength training: effects on muscle strength and size in untrained individuals. *J. Strength Cond. Res*. 33, S159–S166. doi: 10.1519/JSC.0000000000003036

Ferreira, J. A., and Zwinderman, A. H. (2006). On the Benjamini-Hochberg method. *Ann. Stat*. 34, 1827–1849. doi: 10.1214/009053606000 00425

Franchi, M. V., Longo, S., Mallinson, J., Quinlan, J. I., Taylor, T., Greenhaff, P. L., et al. (2018). Muscle thickness correlates to muscle cross-sectional area in the assessment of strength training-induced hypertrophy. *Scand. J. Med. Sci. Sport* 28, 846–853. doi: 10.1111/sms.12961

Fukunaga, T., Roy, R. R., Shellock, F. G., Hodgson, J. A., Day, M. K., Lee, L. P., et al. (1992). Physiological cross-sectional area of human leg muscles based on magnetic resonance imaging. *J. Orthop. Res*. 10, 928–934. doi: 10.1002/jor.1100100623

Giavarina, D. (2015). Understanding Bland Altman analysis lessons in biostatistics. *Biochem. Medica*. 25, 141–151. doi: 10.11613/BM.2015.015

Giles, L. S., Webster, K. E., McClelland, J. A., and Cook, J. (2015). Can ultrasound measurements of muscle thickness be used to measure the size of individal quadriceps muscles in people with patellofemoral pain? *Phys. Ther. Sport* 16, 45–52. doi: 10.1016/j.ptsp.2014.04.002

Grouven, U., Bender, R., Ziegler, A., and Lange, S. (2007). Comparing methods of measurement. *Dtsch. Med. Wochenschr*. 132, 69–73. doi: 10.1055/s-2007-959047

Guthrie, R., Grindstaff, T. L., Croy, T., Ingersoll, C. D., and Saliba, S. A. (2012). The effect of traditional bridging or suspension-exercise bridging on lateral abdominal thickness in individuals with low back pain. *J. Sport Rehabil*. 21, 151–160. doi: 10.1123/jsr.21.2.151

Hebert, J. J., Koppenhaver, S. L., Parent, E. C., and Fritz, J. M. A. (2009). Systematic review of the reliability of rehabilitative ultrasound imaging for the

quantitative assessment of the abdominal and lumbar trunk muscles. *Spine* 34, 848–856. doi: 10.1097/BRS.0b013e3181ae625c

Khoury, V., Cardinal, É., and Brassard, P. (2008). Atrophy and fatty infiltration of the supraspinatus muscle: sonography vs. MRI. *Am. J. Roentgenol.* 190, 1105–1111. doi: 10.2214/AJR.07.2835

Kim, Y. S., Heo, N. Y., and Kim, M. W. (2011). The test-retest reliability of supraspinatus cross-sectional area measurement by sonography. *Ann. Rehabil. Med.* 35, 524. doi: 10.5535/arm.2011.35.4.524

Koch, R., and Spörl, E. (2007). Statistische verfahren zum vergleich zweier messmethoden und zur kalibrierung: konkordanz-, korrelations- und regressionsanalyse am beispiel der augeninnendruckmessung. *Klin. Monbl. Augenheilkd.* 224, 52–57. doi: 10.1055/s-2006-92 7278

König, N., Cassel, M., Intziegianni, K., and Mayer, F. (2014). Inter-rater reliability and measurement error of sonographic muscle architecture assessments. *J. Ultrasound Med.* 33, 769–777. doi: 10.7863/ultra.33.5.769

Koppenhaver, S. L., Parent, E. C., Teyhen, D. S., Hebert, J. J., and Fritz, J. M. (2009). The effect of averaging multiple trials on measurement error during ultrasound imaging of transversus abdominis and lumbar multifidus muscles in individuals with low back pain. *J. Orthop. Sports Phys. Ther.* 39, 604–611. doi: 10.2519/jospt.2009.3088

Kordi, M., Folland, J., Goodall, S., Haralabidis, N., Maden-Wilkinson, T., Sarika Patel, T., et al. (2020). Mechanical and morphological determinants of peak power output in elite cyclists. *Scand. J. Med. Sci. Sport* 30, 227–237. doi: 10.1111/sms.13570

Kwiecien, R., Kopp-Schneider, A., and Blettner, M. (2011). Konkordanzanalyse: teil 16 der serie zur bewertung wissenschaftlicher publikationen. *Dtsch. Arztebl.* 108, 515–521. doi: 10.3238/arztebl.2011.0515

Larivière, C., Henry, S. M., Gagnon, D. H., Preuss, R., and Dumas, J. P. (2019). Ultrasound measures of the abdominal wall in patients with low back pain before and after an 8-week lumbar stabilization exercise program, and their association with clinical outcomes. *PM R.* 11, 710–721. doi: 10.1002/pmrj.12000

Lin, L. I. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 45, 255–268. doi: 10.2307/2532051

Lopes, L. C. C., Mota, J. P., Prestes, J., Schincagalia, R. M., Silva, D. M., Queiroz, N. P., et al. (2019). Intradialytic resistance training improves capacity and lean mass gain in individuals on hemodialysis: a randomized pilot trial. *Arch. Phys. Med. Rehabil.* 100, 2151–2158. doi: 10.1016/j.apmr.2019.06.006

Matos, F., Amaral, J., Martinez, E., Canário-Lemos, R., Moreira, T., Cavalcante, J., et al. (2022). Changes in muscle thickness after 8 weeks of strength training, electromyostimulation and both combined in healthy young adults. *Int. J. Environ. Res. Public Health* 19, 63184. doi: 10.3390/ijerph19063184

May, S., Locke, S., and Kingsley, M. (2021). Gastrocnemius muscle architecture in elite basketballers and cyclists: a cross-sectional cohort study. *Front. Sport Act Living* 3, 768846. doi: 10.3389/fspor.2021.768846

Mayes, S. J., Baird-Colt, P. H., and Cook, J. L. (2015). Ultrasound imaging is a valid method of measuring the cross-sectional area of the quadratus femoris muscle. *J. Dance Med. Sci.* 19, 3–10. doi: 10.12678/1089-313X.19.1.3

McAuliffe, S., Mc Creesh, K., Purtill, H., and O'Sullivan, K. A. (2017). Systematic review of the reliability of diagnostic ultrasound imaging in measuring tendon size: is the error clinically acceptable? *Phys. Ther. Sport* 26, 52–63. doi: 10.1016/j.ptsp.2016.12.002

Mendis, M. D., Wilson, S. J., Stanton, W., and Hides, J. A. (2010). Validity of real-time ultrasound imaging to measure anterior hip muscle size: a comparison with magnetic resonance imaging. *J. Orthop. Sports Phys. Ther.* 40, 577–581. doi: 10.2519/jospt.2010.3286

Nabavi, N., Mosallanezhad, Z., Haghighatkhah, H. R., and Ali Mohseni Bandpeid, M. (2014). Reliability of rehabilitative ultrasonography to measure transverse abdominis and multifidus muscle dimensions. *Iran J. Radiol.* 11, e21008. doi: 10.5812/iranjradiol.21008

Ozaki, H., Sawada, S., Osawa, T., Natsume, T., Yoshihara, T., Deng, P., et al. (2020). Muscle size and strength of the lower body in supervised and in combined supervised and unsupervised low-load resistance training. *J. Sport Sci. Med.* 19, 721–726.

Padulo, J., Trajković, N., Cular, D., Grgantov, Z., Madić, D. M., Vico, D. R., et al. (2020). Validity and reliability of isometric-bench for knee isometric assessment. *Int J Environ. Res. Public Health* 17, 1–8. doi: 10.3390/ijerph17124326

Palmer, T. P., Akehi, K., Thiele, R. M., Smith, D. B., and Thompson, J. (2015). Reliability of panoramic ultrasound imaging in simultaneously examining muscle size and quality of the hamstric muscles in young, healthy males and females. *Ultrasound Med. Biol.* 41, 675–684. doi: 10.1016/j.ultrasmedbio.2014.10.011

Panidi, I., Bogdanis, G. C., Terzis, G., Donti, A., Konrad, A., Gaspari, V., et al. (2021). Muscle architectural and functional adaptations following 12-weeks of stretching in adolescent female athletes. *Front. Physiol.* 12, 701338. doi: 10.3389/fphys.2021.701338

Perkisas, S., Bastijns, S., Stéphane, B., Bauer, J., Beaudart, C., David, B., et al. (2021). Application of ultrasound for muscle assessment in sarcopenia: 2020 SARCUS update. *Eur. Geriatr. Med.* 12, 45–59. doi: 10.1007/s41999-020-00433-9

Prestes, J., Tibiana, R. A., Araujo Sousa de, E., da ConhaNascimento, D., de oliveira Rocha, P., Camarco, N., et al. (2019). Strength and Muscular Adaptations After 6 Weeks of Rest-Pause vs. traditional multiple-sets resistance training in trained subjects. *J. Strength Cond. Res.* 33,S113–S121. doi: 10.1519/JSC.0000000000001923

Rahmani, N., Karimian, A., Mohseni-Bandpei, M. A., and Bassampour, S. A. (2019). Reliability of sonography in the assessment of lumbar stabilizer muscles size in healthy subjects and patients with scoliosis. *J. Bodyw. Mov. Ther.* 23, 138–141. doi: 10.1016/j.jbmt.2018.05.010

Raj, I. S., Bird, S. R., and Shield, A. J. (2012). Reliability of ultrasonographic measurement of the architecture of the vastus lateralis and gastrocnemius medialis muscles in older adults. *Clin. Physiol. Funct. Imaging* 32, 65–70. doi: 10.1111/j.1475-097X.2011.01056.x

Rosenberg, J. G., Ryan, E. D., Sobolewski, E. J., Scharville, M. J., Thompson, B. J., King, G. E., et al. (2014). Reliability of panoramic ultrasound imaging to simultaneously examine muscle size and quality of the medial gastrocnemius. *Muscle Nerve* 49, 736–740. doi: 10.1002/mus.24061

Ruple, B. A., Smith, M. A., Osburn, S. C., Sexton, C. L., Dowin, J. S., Edison, J. L., et al. (2022). Compariosns between skeletal muscle imaging techniques and histology in tracking midthigh hypertrophic adaptations following 10 wk of resistance training. *J. Appl. Physiol.* 133, 416–425. doi: 10.1152/japplphysiol.00219.2022

Rustani, K., Kundisova, L., Capecchi, P. L., Nante, N., and Bicchi, M. (2019). Ultrasound measurement of rectus femoris muscle thickness as a quick screening test for sarcopenia assessment. *Arch. Gerontol. Geriatr.* 83, 151–154. doi: 10.1016/j.archger.2019.03.021

Sarto, F., Spörri, J., Fitze, D. P., Quinlan, J. I., Narici, M. V., Franchi, M. V., et al. (2021). Implementing ultrasound imaging for the assessment of muscle and tendon properties in elite sports: practical aspects, methodological considerations and future directions. *Sport Med.* 51, 1151–1170. doi: 10.1007/s40279-021-01436-7

Schober, P., and Schwarte, L. A. (2018). Correlation coefficients: appropriate use and interpretation. *Anesth. Analg.* 126, 1763–1768. doi: 10.1213/ANE.0000000000002864

Schoenfeld, B. J., Contreras, B., Krieger, J., Grgic, J., Delcastillo, K., Belliard, R., et al. (2019). Resistance training volume enhances muscle hypertrophy but not strength in trained men. *Med. Sci. Sports Exerc.* 51, 94–103. doi: 10.1249/MSS.0000000000001764

Schoenfeld, B. J., Ogborn, D., and Krieger, J. W. (2017). The dose-response relationship between resistance training volume and muscle hypertrophy: are there really still any doubts? *J. Sport Sci.* 35, 1985–1987. doi: 10.1080/02640414.2016.1243800

Schoenfeld, B. J., Pope, Z. K., Benik, F. M., Hester, G. M., Sellers, J., Nooner, J. L., et al. (2016). Longer interset rest periods enhance muscle strength and hypertrophy in resistance-trained men. *J. Strength Cond. Res.* 30, 1805–1812. doi: 10.1519/JSC.0000000000001272

Scott, J. M., Martin, D. S., Ploutz-snyder, R., Matz, T., Caine, T., Downs, M., et al. (2017). Panoramic ultrasound : a novel and valid tool for monitoring change in muscle mass. *J. Cachexia. Sarcopenia Muscle* 8, 475–481. doi: 10.1002/jcsm.12172

Simpson, C. L., Kim, B. D. H., Bourcet, M. R., Jones, G. R., and Jakobi, J. M. (2017). Stretch training induces unequal adaptation in muscle fascicles and thickness in medial and lateral gastrocnemii. *Scand. J. Med. Sci. Sport* 27, 1597–1604. doi: 10.1111/sms.12822

Souza, E. O., Ugrinowitsch, C., Tricoli, V., Roschel, H., Lowery, R. P., Aihara, A. Y., et al. (2014). Early adaptations to six weeks of non-periodized and periodized strength training regimens in recreational males. *J. Sport Sci. Med.* 13, 604–609.

Tavares, L. D., de Souza, E. O., Urinowitsch, C., Laurentino, G. C., Roschel, H., Aihara, A. Y., et al. (2017). Effects of different strength training frequencies during reduced training period on strength and muscle cross-sectional area. *Eur. J. Sport Sci.* 17, 665–672. doi: 10.1080/17461391.2017.1298673

Temes, A., Clifton, A. T., Hilton, V., Girard, L., Strait, N., Karduna, A., et al. (2014). Reliability and validity of thickness measurements of the supraspinatus muscle of the schoulder: an ultrasonography study. *J. Sport Rehabil.* 23, 2013–0023. doi: 10.1123/jsr.2013-0023

Teyhen, D., and Koppenhaver, S. (2011). Rehabilitative ultrasound imaging. *J. Physiother.* 57, 196. doi: 10.1016/S1836-9553(11)70044-3

Thomaes, T., Thomis, M., Onkelinx, S., Coudyzer, W., and Vanhess, V. C. (2012). Reliability and validity of ultrasound technique to measure the rectus femoris muscle diameter in older CAD-patients. *BMC Medi. Image* 12, 7. doi: 10.1186/1471-2342-12-7

Ticinesi, A., Meschi, T., Narici, M. V., Lauretani, F., and Maggio, M. (2017). Muscle ultrasound and sarcopneia in older individuals: a clinical perspective. *J. Am. Med. Dir. Assoc.* 18, 290–300. doi: 10.1016/j.jamda.2016.11.013

Ticinesi, A., Narici, M. V., Lauretani, F., Nouvenne, A., Colizzi, E., Mantovani, M., et al. (2018). Assessing sarcopenia with vastus lateralis muscle ultrasound: an operative protocol. *Aging Clin. Exp. Res.* 30, 1437–43. doi: 10.1007/s40520-018-0958-1

Vikberg, S., Sörlén, N., Brandén, L., Johansson, J., Nordström, A., Hult, A., et al. (2019). Effects of resistance training on functional strength and muscle mass in 70-year-old individuals with pre-sarcopenia: a randomized controlled trial. *J. Am. Med. Dir. Assoc.* 20, 28–34. doi: 10.1016/j.jamda.2018.09.011

Wackerhage, H., Schoenfeld, B. J., Hamilton, D. L., Lehti, M., and Hulmi, J. J. (2019). Stimuli and sensors that initiate muscle hypertrophy following resistance exercise. *J. Appl. Physiol.* 126, 30–43. doi: 10.1152/japplphysiol.00685.2018

Wada, T., Tanishima, S., Kitsuda, Y., Osaki, M., Nagashima, H., Hagino, H., et al. (2020). Preoperative low muscle mass is a predictor of falls within 12 months of surgery in patients with lumbar spinal stenosis. *BMC Geriatr.* 20, 1–8. doi: 10.1186/s12877-020-01915-y

Wallwork, T. L., Hides, J. A., and Stanton, W. R. (2007). Intrarater and interrater reliability of assessment of lumbar multifidus muscle thickness using rehabilitative ultrasound imaging. *J. Orthop. Sports Phys. Ther.* 37, 608–612. doi: 10.2519/jospt.2007.2418

Wang, Y., Ikeda, S., and Ikoma, K. (2021). Passive repetitive stretching is associated with greater muscle mass and cross-sectional area in the sarcopenic muscle. *Sci. Rep.* 11, 15302. doi: 10.1038/s41598-021-94709-0

Warneke, K., Brinkmann, A., Hillebrecht, M., and Schiemann, S. (2022). Influence of long-lasting static stretching on maximal strength, muscle thickness and flexibility. *Front. Physiol.* 13, 878955. doi: 10.3389/fphys.2022.878955

Watanabe, Y., Madarame, H., Ogasawara, R., Nakazato, K., and Ishii, N. (2014). Effect of very low-intensity resistance training with slow movement on muscle size and strength in healthy older adults. *Clin. Physiol. Funct. Imaging* 34, 463–470. doi: 10.1111/cpf.12117

Wirth, K., Atzor, K. R., and Schmidtbleicher, D. (2007). Veränderungen der muskelmasse in abhängigkeit von trainingshäufigkeit und leistungsniveau. *Dtsch. Z Sportmed.* 58, 178–183.

Wong, A. Y. L., Parent, E., and Kawchuk, G. (2013). Reliability of 2 ultrasonic imaging analysis methods in quantifying lumbar multifidus thickness. *J. Orthop. Sports Phys. Ther.* 43, 251–262. doi: 10.2519/jospt.2013.4478

Worsley, P. R., Kitsell, F., Samuel, D., and Stokes, M. (2014). Validity of measuring distal vastus medialis muscle using rehabilitative ultrasound imaging vs. magnetic resonance imaging. *Man. Ther.* 19, 259–263. doi: 10.1016/j.math.2014.02.002

Yahata, K., Konrad, A., Sato, S., Kiyono, R., Yoshida, R., Fukaya, T., et al. (2021). Effects of a high-volume static stretching programme on plantar-flexor muscle strength and architecture. *Eur. J. Appl. Physiol.* 121, 1159–1166. doi: 10.1007/s00421-021-04608-5

Yi, T. I., Han, I. S., Kim, J. S., Jin, J. R., and Han, J. S. (2012). Reliability of the supraspinatus muscle thickness measurement by ultrasonography. *Ann. Rehabil. Med.* 36, 488–495. doi: 10.5535/arm.2012.36.4.488

Zaras, N., Stasinaki, A. N., and Terzis, G. (2021). Biological determinants of track and field throwing performance. *J. Funct. Morphol. Kinesiol.* 6, 40 doi: 10.3390/jfmk6020040