Rev Bras Cineantropom Desempenho Hum

review article

https://doi.org/10.1590/1980-0037.2023v25e93936

Validity, reliability and measurement error of quadriceps femoris muscle thickness obtained by ultrasound in healthy adults: a systematic review

Validade, confiabilidade e erro da medida da espessura muscular do quadríceps femoral obtida pela ultrassonografia em adultos saudáveis: uma revisão sistemática

André Luiz Conveniente Soares[1]
https ://orcid.org/0000-0003-0536-4164
Ramon Franco Carvalho[1]
https ://orcid.org/0000-0003-3093-582X
Roberto Mogami[2]
https ://orcid.org/0000-0002-7610-2404
Cláudia de Mello Meirelles[3]
https ://orcid.org/0000-0003-0673-370X
Paulo Sergio Chagas Gomes[1]
https ://orcid.org/0000-0002-1026-8639

1 Universidade do Estado do Rio de Janeiro. Instituto de Educação Física e Desportos. Laboratório Crossbridges. Rio de Janeiro, RJ. Brasil.
2 Universidade do Estado do Rio de Janeiro. Faculdade de Ciências Médicas. Rio de Janeiro, RJ. Brasil.
3 Escola de Educação Física do Exército. Setor de Pesquisa e Extensão. Rio de Janeiro, RJ. Brasil.

Received: April 20, 2023
Accepted: July 11, 2023

How to cite this article
Soares ALC, Carvalho RF, Mogami R, Meirelles CM, Gomes PSC. Validity, reliability and measurement error of quadriceps femoris muscle thickness obtained by ultrasound in healthy adults: a systematic review. Rev Bras Cineantropom Desempenho Hum 2023, 25:e93936. DOI: https://doi.org/10.1590/1980-0037.2023v25e93936

Corresponding author
Paulo Sergio Chagas Gomes.
Laboratório Crossbridges, Programa de Pós-graduação em Ciências do Exercício e do Esporte, Instituto de Educação Física e Desportos, Universidade do Estado do Rio de Janeiro
Rua São Francisco Xavier, n. 524, Bloco F, 8º andar, Sala 8104, 20550-900, Maracanã, Rio de Janeiro (RJ), Brasil.
E-mail: paulo.gomes@uerj.br

Abstract - Due to its low cost and operational simplicity, ultrasound has been used to monitor muscle thickness in laboratory environments, rehabilitation clinics, and sports clubs. However, it is necessary to determine the measurement's quality to infer whether the possible changes observed are derived from the treatment or the measurement error. Therefore, we performed a systematic review to determine the validity, reliability, and measurement error of quadriceps femoris muscle thickness obtained by ultrasound in healthy adults. A search was conducted in the Pubmed, Scopus, and Web of Science databases until April 2022. The study selection process was carried out by two independent researchers, with the presence of a third researcher in case of disagreements. Twenty-six studies were eligible for the review, being 4 of validity, 4 of reliability only, and 18 of reliability and measurement error. The intraclass correlation coefficient ranged from 0.60 to 0.99 in validity studies and from 0.44 to 0.99 in reliability studies. The typical error of measurement ranged from 0.01 to 0.47 cm, and the coefficient of variation was from 0.5 to 17.9%. Four studies received "very good" classification in all the risk of bias analysis criteria. Therefore, it is concluded that the quadriceps femoris muscle thickness obtained by ultrasound was shown to be valid, reliable, and to have low measurement errors in healthy adults. The weighted average of the relative error was 6.5%, less than typical increases in resistance training studies. The raters' experience and methodological care for repeated measurements were necessary to observe low measurement errors.

Key words: Atrophy; Hypertrophy; Magnetic resonance imaging; Reproducibility; Ultrasonography.

Resumo - *Devido ao baixo custo e simplicidade operacional, a ultrassonografia tem sido utilizada para monitorar a espessura muscular em ambientes laboratoriais, clínicas de reabilitação e clubes desportivos. Porém, é necessário determinar a qualidade da medida para inferir se as possíveis modificações observadas são decorrentes do tratamento ou do erro da medida. Portanto, realizamos uma revisão sistemática para determinar a validade, confiabilidade e erro da medida da espessura muscular do quadríceps femoral obtida pela ultrassonografia em adultos saudáveis. Foi realizada busca nas bases de dados Pubmed, Scopus e Web of Science até abril de 2022. O processo de seleção dos estudos foi realizado por dois pesquisadores independentes, com a presença de um terceiro pesquisador em caso de divergências. Vinte e seis estudos foram elegíveis para a revisão, sendo 4 de validade, 4 apenas de confiabilidade e 18 de confiabilidade e erro de medida. O coeficiente de correlação intraclasse variou de 0,60 a 0,99 em estudos de validade e 0,44 a 0,99 em estudos de confiabilidade. O erro típico da medida variou de 0,01 a 0,47 cm e o coeficiente de variação foi de 0,5 a 17,9%. Quatro estudos receberam classificação "muito bom" em todos os critérios na análise de risco de viés. Portanto, concluiu-se que a espessura muscular do quadríceps femoral obtida pela ultrassonografia se mostrou válida, confiável e com baixos erros de medida em adultos saudáveis. A média ponderada do erro relativo foi de 6,5%, menor do que os aumentos típicos em estudos de treinamento resistido. A experiência dos avaliadores e o cuidado metodológico com as medidas repetidas foram necessários para observar baixos erros de medida.*

Palavras-chave: *Atrofia; Hipertrofia; Imagem por ressonância magnética; Reprodutibilidade dos testes; Ultrassom.*

INTRODUCTION

Muscle thickness (MT) obtained through ultrasound (US) has been used to monitor hypertrophy[1,2] and muscle atrophy[3,4] in the quadriceps femoris. The US uses waves with varying frequencies that penetrate the body while traveling through tissues with different acoustic impedances and reflecting echoes to the transducer, which are converted into electrical signals[5]. The angle and pressure of the transducer on the skin interfere with the measurement, as the incorrect positioning of the transducer can cause the reflected echoes not to be detected[6,7].

For a quality image, there is a need for more outstanding care in positioning the transducer based on a more detailed methodological description[8,9], allowing the records made with the US to be replicated when there is a need to perform repeated measurements[10,11]. This need is essential in experimental studies when treatment is applied to the muscle tissue, such as resistance training, where small changes in MT are often observed[12,13].

US is a commonly used to measure muscle architecture variables, such as quadriceps femoris MT[14,15]. Its operational simplicity, low cost compared to magnetic resonance imaging (MRI) or computed tomography (CT), and ease of image evaluation with free software make it attractive in research laboratory environments, rehabilitation clinics, and sports clubs. Therefore, it is necessary to verify valid and reliable ultrasound methods capable of monitoring quadriceps femoris MT in the literature. It is also necessary to verify the magnitude of the measurement error in order to be able to infer whether the possible changes observed experimentally are derived from the treatment itself instead of caused by measurement error.

The study aimed to determine the validity, reliability, and magnitude of measurement error of MT of the rectus femoris, vastus lateralis, vastus medialis, and vastus intermedius muscles obtained by the US in healthy adults.

METHODS

Protocol and registration

This systematic review followed the recommendations of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statement[16]. It was registered in the International Prospective Register of Systematic Reviews (PROSPERO) under the identification CRD42020205566.

Eligibility criteria

US studies that performed a validity test comparing the measurements of MT in cadavers or in vivo through MRI or CT could be included. Studies that tested the relative reliability or error of intra- or inter-rater measurement of MT for healthy adults aged 18 to 65 were also included. The muscles observed here were: rectus femoris, vastus lateralis, vastus medialis, and vastus intermedius. Studies in English and Portuguese were reviewed. Abstracts published in conference proceedings, dissertations, theses, inadequate measures or analyses, literature review studies, and research reports were excluded.

Search strategy

Searches were performed in Pubmed, Scopus, and Web of Science databases until April 2022. The following terms were combined: validity, reliability, measurement error, error of measurement, coefficient of variation, thickness, quadriceps femoris, rectus femoris, vastus lateralis, vastus medialis, and vastus intermedius. The terms were combined using the Boolean operators "AND" between descriptors and "OR" between tests and muscles. The search equation was adjusted to the specificity of each database. A manual search was also performed in references cited in published studies on similar topics.

Study selection

After removing the duplicates, there was the screening process, where the title and abstract of the studies were read by two researchers independently. In cases of conflicting screening, a discussion between the researchers was carried out to keep the article in the review. When disagreement occurred, a third researcher made the final decision. Subsequently, the same researchers read potentially eligible articles to select studies that met the eligibility criteria. Again, in cases of disagreement, a third researcher evaluated the studies and determined their permanence or exclusion from the review.

Risk of bias

Two researchers performed the risk of bias analysis independently. When there was disagreement, the researchers discussed the difference. A third researcher made the final decision when there was no consensus. The risk of bias in the validity, reliability, and measurement error studies was analyzed according to the Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN)[17]. Seven criteria were evaluated, classifying them into five different discriminatory states.

Ratings were as follows: very good, when there was convincing evidence or arguments provided that the standard was met; adequate, when it is assumed, although not explicitly described, that the standard has been met; doubtful when it was unclear whether the standard was met; inadequate, when there was evidence that the standard was not met; information not available when there was no information to help in the judgment of the criterion.

Data extraction

One researcher extracted data from the studies that the second researcher later checked. The following data were extracted: n sample, gender and age of the participants, validity test (cadaver, magnetic resonance imaging, or computed tomography), type of reliability or measurement error (intra- or inter-rater), muscles (rectus femoris, vastus lateralis, vastus medialis or vastus intermedius) and statistical indices, such as intraclass correlation coefficient (ICC), typical error of measurement (TEM), standard error of measurement (SEM) and coefficient of variation (CV).

Weighted average of relative error

The statistical index's weighted average (WA) representing the relative error (TEM%, SEM%, or CV%) was performed, considering sample n, according to the equation below. The highest reported error value was considered when the result was presented through amplitude bands with the lowest and highest error value.

$$WA = \frac{\sum(RE \times n)}{\sum n} \quad (1)$$

Where:
WA = weighted average
Σ = sum
RE = relative error
n = number of subjects

RESULTS

Study selection

The search identified 375 records, 101 in the PubMed database, 114 in Scopus, and 160 in the Web of Science. Three records from other sources were added (studies detected from the reference of other studies). One hundred thirty-one duplicates were removed, and 247 records were selected for screening. After reading the title and abstract, 211 records were excluded, and 36 articles were selected for eligibility. Subsequently, the full text was read, and ten studies were excluded for different reasons. Six studies presented inadequate samples, two did not measure MT, one did not perform an adequate analysis, and one did not inform the type of comparison. The summary of the selection of studies is presented in Figure 1 in the form of a flowchart.

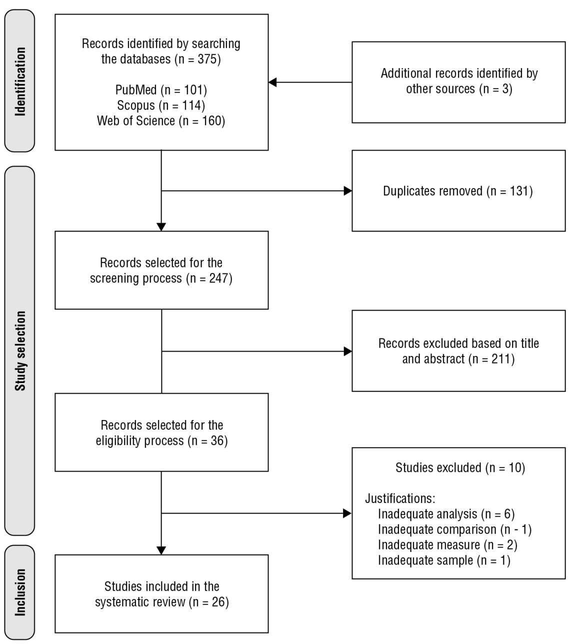

Figure 1. Study selection flowchart.

Studies characteristics

Of the 26 studies included, 4 were validity studies[18-21], and 4 were reliability only[22-25]. Eighteen were of reliability and measurement error[26-43]. Four hundred ninety subjects were included, from which 311 men, 164 women, and 15 were not identified. Participants were primarily young adults, and the mean age ranged from 18 to 49. The characteristics of the included studies are summarized in Table 1.

Table 1. Overview of eligible studies.

Studies	N Gender	Age (years)	Study type	Comparison	Muscle	Statistical index
Arruda et al.[26]	25M	24±4	Reliability & measurement error	Intra- & inter-rater	VL	Intra: ICC = 0.991-0.998; TEM = 0.01-0.03 cm; CV = 0.5-1.8% Inter: ICC = 0.990-0.996; TEM = 0.02-0.03 cm; CV = 0.6-2.3%
Barotsis et al.[22]	8M & 5W	24±3	Reliability	Intra-rater	RF VI	ICC = 0.621-0.976 ICC = 0.411-0.938
Betz et al.[18]	7M & 9W	33±11 47±11	Validity & Reliability	MRI, intra- & inter-rater	VL	Validity: ICC = 0.835-0.895 Intra: ICC = 0.928-0.961 Inter: ICC = 0.936-0.965
Caresio et al.[23]	25M & 25W	31±11	Reliability	Inter-rater	RF VL	ICC = 0.98 ICC = 0.99
Carr et al.[27]	7M & 10W	24±3 21±2	Reliability & measurement error	Inter-rater	VL	ICC = 0.826-0.854; SEM = 0.19 cm; SEM% = 7.7-7.8%
Chiaramonte et al.[24]	21M	26-38	Reliability	Intra- & inter-rater	VL	Intra: ICC = 0.92 Inter: ICC = 0.97
Cleary et al.[28]	15 NI	18-35	Reliability & measurement error	Intra- & inter-rater	RF VL	Intra: SEM = 0.03-0.07 cm; SEM% = 1.2-2.8% Inter: ICC = 0.984; SEM = 0.06 cm Intra: SEM = 0.04-0.08 cm; SEM% = 1.6-3.3% Inter: ICC = 0.993; SEM = 0.05 cm
Dudley-Javoroski et al.[29]	8M & 8W	26±5	Reliability & measurement error	Intra- & inter-rater	VL	Intra: ICC = 0.66-0.99; CV = 0.8-6.4% Inter: ICC = 0.742
Ema et al.[30]	7M & 7W	24±1	Reliability & measurement error	Intra-rater	RF	ICC = 0.981-0.984; CV = 2.3-2.4%
Franchi et al.[31]	9M	24±2	Reliability & measurement error	Intra-rater	VL	ICC = 0.99; SEM% = 1.7%
Gomes et al.[32]	7M & 8W	34±11	Reliability & measurement error	Intra-rater	RF	ICC = 0.929; CV = 4.6%
Hagoort et al.[33]	7M & 5W	23±4	Reliability & measurement error	Intra- & inter-rater	VL	Intra: ICC = 0.93-0.98; SEM = 0.05-0.11 cm Inter: ICC = 0.98; SEM = 0.10 cm
Ishida et al.[34]	14M	21±1	Reliability & measurement error	Intra- & inter-rater	RF	Intra: ICC = 0.99; SEM = 0.04 cm Inter: ICC = 0.96; SEM = 0.07 cm
Jacob et al.[35]	32M	18±1	Reliability & measurement error	Intra-rater	VL	ICC = 0.95; SEM = 0.04 cm

Note. CV = coefficient of variation; ICC = intraclass correlation coefficient; M = men; MRI = magnetic resonance imaging; N = number of subjects; NI = not informed; RF = rectus femoris; SEM = standard error of measurement; TEM = typical error of measurement; VI = vastus intermedius; VL = vastus lateralis; VM = vastus medialis; W = women.

Table 1. Continued...

Studies	N Gender	Age (years)	Study type	Comparison	Muscle	Statistical index
Lanferdini et al.[36]	11M & 11W	27±5	Reliability & measurement error	Intra- & inter-rater	RF	Intra: ICC = 0.848; SEM = 0.11 cm; CV = 14.8%
						Inter: ICC = 0.803; SEM = 0.15 cm; CV = 17.9%
					VL	Intra: ICC = 0.987; SEM = 0.04 cm; CV = 14.7%
						Inter: ICC = 0.882; SEM = 0.09 cm; CV = 13.1%
Lima and Oliveira[37]	4M & 10W	22±2	Reliability & measurement error	Intra-rater	VL	ICC = 0.95-0.99; TEM = 0.10-0.11 cm; CV = 3.1-3.8%
Mairet et al.[38]	10M & 9W	30±6	Reliability & measurement error	Intra-rater	VL	ICC = 0.65-0.75; SEM = 0.15-0.18; SEM% = 6.7-7.9%
Mechelli et al.[19]	10M & 10W	49±10	Validity	MRI	RF + VI	ICC = 0.99; SEM = 0.07 cm
Mechelli et al.[39]	12M & 12W	49±10	Reliability & measurement error	Intra- & inter-rater	RF + VI	Intra: ICC = 0.96; SEM = 0.13 cm
						Inter: ICC = 0.98; SEM = 0.10 cm
Nijholt et al.[20]	5M & 9W	33±NI	Validity & Reliability	MRI & intra-rater	RF	Validity: ICC = 0.60
						Reliability: ICC = 0.87
Oranchuck et al.[40]	26M	29±5	Reliability & measurement error	Intra-rater	RF	ICC = 0.93-0.95; TEM = 0.23-0.47 cm; CV = 2.7-4.1%
					VL	ICC = 0.94-0.98; TEM = 0.15-0.26 cm; CV = 2.4-3.8%
					VI	ICC = 0.88-0.98; TEM = 0.14-0.37 cm; CV = 2.8-9.3%
Ruas et al.[41]	10M	23±2	Reliability & measurement error	Intra-rater	RF	ICC = 0.97; SEM = 0.07 cm
					VL	ICC = 0.97; SEM = 0.10 cm
					VM	ICC = 0.97; SEM = 0.14 cm
					VI	ICC = 0.99; SEM = 0.07 cm
Santos and Armada-da-Silva[42]	10M & 10W	20±2	Reliability & measurement error	Intra-rater	RF	ICC = 0.88-0.99; SEM = 0.09-0.14 cm
					VL	ICC = 0.70-0.99; SEM = 0.09-0.15 cm
					VM	ICC = 0.80-0.98; SEM = 0.07-0.16 cm
					VI	ICC = 0.74-0.99; SEM = 0.13-0.19 cm
Soares et al.[43]	12M	24±6	Reliability & measurement error	Intra-rater	VL	ICC = 0.964; TEM = 0.07 cm; CV = 2.9%
Takahashi et al.[25]	12M	27±4	Reliability	Intra- & inter-rater	RF	Intra: ICC = 0.95
						Inter: ICC = 0.70
Worsley et al.[21]	12M	18-30	Validity & Reliability	MRI & intra-rater	VM	Validity: ICC = 0.84-0.94
						Reliability: ICC = 0.90-0.98

Note. CV = coefficient of variation; ICC = intraclass correlation coefficient; M = men; MRI = magnetic resonance imaging; N = number of subjects; NI = not informed; RF = rectus femoris; SEM = standard error of measurement; TEM = typical error of measurement; VI = vastus intermedius; VL = vastus lateralis; VM = vastus medialis; W = women.

Risk of bias in studies

The risk of bias analysis performed using the COSMIN tool showed that four studies were classified as "very good" in the seven criteria[20,21,26,39] and two with at least one "inadequate" criterion[22,27]. Table 2 presents the classification of studies for each of the seven criteria.

Table 2. Analysis of the risk of bias by the COSMIN tool.

Studies	C1	C2	C3	C4	C5	C6	C7
Arruda et al.[26]	Very good	Very good	Very good	Very good	Very good	Very good	Very good
Barotsis et al.[22]	Inadequate	Very good	Adequate	Very good	Doubtful	Very good	Very good
Betz et al.[18]	Very good	Very good	Adequate	Adequate	Very good	Very good	Adequate
Caresio et al.[23]	Very good	Very good	Adequate	Very good	Very good	Very good	Very good
Carr et al.[27]	Very good	Very good	Adequate	Inadequate	Very good	Very good	Very good
Chiaramonte et al.[24]	Very good	Very good	Adequate	Adequate	Adequate	Very good	Adequate
Cleary et al.[28]	Very good	Very good	Adequate	Adequate	Very good	Very good	Very good
Dudley-Javoroski et al.[29]	Adequate	Doubtful	Adequate	Adequate	Adequate	Doubtful	Very good
Ema et al.[30]	Very good	Very good	Very good	Adequate	Adequate	Very good	Adequate
Franchi et al.[31]	Adequate	Doubtful	Adequate	Adequate	Doubtful	Very good	Very good
Gomes et al.[32]	Doubtful	Very good	Very good	Adequate	Very good	Very good	Adequate
Hagoort et al.[33]	Adequate	Very good	Adequate	Adequate	Very good	Very good	Very good
Ishida et al.[34]	Very good	Very good	Very good	Very good	Very good	Very good	Adequate
Jacob et al.[35]	Very good	Very good	Adequate	Adequate	Adequate	Very good	Very good
Lanferdini et al.[36]	Very good	Very good	Adequate	Adequate	Adequate	Very good	Adequate
Lima and Oliveira[37]	Very good	Very good	Very good	Doubtful	Doubtful	Very good	Adequate
Mairet et al.[38]	Adequate	Adequate	Adequate	Adequate	Adequate	Doubtful	Adequate
Mechelli et al.[19]	Very good	Very good	Very good	Very good	Very good	Very good	Very good
Mechelli et al.[39]	Very good	Very good	Adequate	Adequate	Doubtful	Very good	Adequate
Nijholt et al.[20]	Very good	Very good	Very good	Very good	Very good	Very good	Very good
Oranchuck et al.[40]	Very good	Very good	Very good	Adequate	Doubtful	Very good	Adequate
Ruas et al.[41]	Very good	Very good	Very good	Adequate	Adequate	Very good	Very good
Santos and Armada-da-Silva[42]	Very good	Very good	Adequate	Adequate	Doubtful	Very good	Very good
Soares et al.[43]	Very good	Very good	Adequate	Adequate	Adequate	Very good	Very good
Takahashi et al.[25]	Very good	Very good	Adequate	Very good	Very good	Very good	Very good
Worsley et al.[21]	Very good	Very good	Very good	Very good	Very good	Very good	Very good

Note. C1 = volunteers were stable in time between repeated measurements; C2 = the time interval between repeated measurements was adequate; C3 = the conditions of repeated measures were similar; C4 = the collection was repeated without knowledge of the values of the previous measurement; C5 = the score values were determined without knowledge of the previous values; C6 = there was some other major flaw in the study design or statistical methods; C7 = the appropriate statistical index for the study was calculated.

Weighted average result

For the WA calculation, the 16 relative error values of the MT of different quadriceps femoris muscles obtained from 12 of the 26 included studies were considered. From the relative error, weighted by the sample n, the WA was 6.5%.

DISCUSSION

The studies included in the systematic review showed that US is valid and reliable for measuring quadriceps femoris MT in healthy adults and having a low absolute and relative measurement error, both intra- and inter-raters. However, for the measurement to be reproducible, the raters must pay attention to the description of the method they will use, including the definition of the measurement location[8], anatomical landmarks[26], stability of the subject[35], positioning of the transducer[34], and experience in image analysis[28].

Four eligible validity studies compared MT obtained by the US versus MRI. Worsley et al.[21] evaluated MT in three different portions of the vastus medialis and observed a high ICC ranging from 0.84 to 0.94. Nijholt et al.[20] observed a moderate ICC of 0.60 in the rectus femoris muscle. However, Mechelli et al.[19] found an almost perfect correlation of 0.99 in the rectus femoris and vastus intermedius muscles. Finally, Betz et al.[18] measured the proximal, medial, and distal portions of the vastus lateralis muscle and observed correlations ranging from 0.835 to 0.895. Other validity studies were not eligible as they were performed with a sample with some disease[44] or with measurements performed in muscle groups other than the quadriceps femoris[45-47]. All the studies mentioned concluded that the US measurement was valid for measuring MT.

Twenty-five studies included in the review performed relative reliability analysis. The lowest ICC value was 0.441[22], and the highest was 0.99, observed in 9 studies[23,26,28,29,31,34,37,41,42]. The low reliability found in the study by Barotsis et al.[22] may have occurred because the MT measurement was performed four times during 24 hours to observe the measurement's reproducibility throughout the day. Participants were instructed to maintain their usual routine in the intervals between collections, including the practice of physical activity, thus impairing stability between measurements.

Higher correlations were observed when the comparison was intra-rater, probably due to the reproducibility of the technique. However, some inter-rater reliability studies have found high correlations when comparing experienced raters against novice raters[27,28,36]. They observed ICC values between 0.803 and 0.993 in rectus femoris and vastus lateralis MT. Cleary et al.[28] suggest that inexperienced and more experienced raters continue to practice their measurements on control images to maintain a high level of reliability before conducting an experimental study. Furthermore, Carr et al.[27] highlighted the need for a detailed method description so that different raters can replicate the technique in different environments and samples.

Although the ICC is a widely used statistical analysis to verify reliability, its results are affected by the heterogeneity of the sample. Thus, it must be accompanied by other analyzes to detect the measurement error, such as TEM or SEM[48,49]. The present review found that absolute errors ranged from 0.01 to 0.47 cm. Our laboratory experience indicates that the methodological care of the entire process, associated with the constant training of the raters, has decreased TEM. In the first study, our group presented an intra-rater TEM of 0.07 cm for vastus lateralis MT[43]. In a recent study, intra- and inter-rater TEM decreased to 0.01 to 0.03 cm for the same variable[26].

In the exercise and sports sciences areas, it was recommended as a criterion that the error of the acceptable relative measure should be at most 10%[48]. Except for one study eligible for this review, all had CV below 10%. Lanferdini et al.[36] observed CVs of 13.1 to 17.9%. The authors argued that the magnitude of the error was probably due to the raters' inexperience with the US measurement.

The WA analysis of the relative error found was 6.5%. This value is a less arbitrary and evidence-based way to define a reasonable cut-off point for the measurement error of quadriceps femoris MT in healthy adults. Previous studies show that it is possible to achieve this index when raters are trained to collect and analyze the measure[26,28,29,31,32,37,43].

Based on the recent experience of our laboratory, it is suggested that the responsible raters carry out a reliability and measurement check before an experimental study, where the US will be used to detect changes in MT. In addition to the precise definition of the measurement site and the training of raters in carrying out the measurement itself, it is recommended the operational description of some procedures based on COSMIN[17], such as: guiding volunteers not to perform physical activity for at least 24 hours before the collection of images, inform the interval between repeated measurements, describe in detail where the transducer will be positioned on the skin to obtain the image of the muscle, encode and shuffle the images in order to blind the raters of the images, experience in the analysis of MT by the software and perform the appropriate statistical analysis to the objectives.

CONCLUSIONS

The current systematic review concluded that the MT of the rectus femoris, vastus lateralis, vastus medialis, and vastus intermedius obtained by the US is a valid, reliable measurement and had low measurement errors in healthy adults. High correlation values were observed for both validity studies and reliability studies. In addition, a low magnitude of measurement errors was observed, with an average error of 6.5%. Experience and care are needed in the steps discussed here to observe low measurement errors.

COMPLIANCE WITH ETHICAL STANDARDS

Funding

Soares ALC was supported with scholarship from Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Capes (protocol number: 88887.603070/2021-00). Gomes PSC was supported by PROCIÊNCIA grant from Universidade do Estado do Rio de Janeiro - UERJ.

Ethical approval

Does not apply to the current study.

Conflict of interest statement

The authors have no conflict of interests to declare.

Author Contributions

Conceived and designed the review: ALCS, RFC, RM, CMM, PSCG; Screened articles: ALCS, RFC, PSCG; Analyzed the data: ALCS, RFC; Wrote the paper: ALCS, PSCG; Critical analysis: RFC, RM, CMM.

REFERENCES

1. Brandner CR, Clarkson MJ, Kidgell DJ, Warmington SA. Muscular adaptations to whole body blood flow restriction training and detraining. Front Physiol 2019;10:1099. http://dx.doi.org/10.3389/fphys.2019.01099. PMid:31551800.

2. Botton CE, Radaelli R, Wilhelm EN, Rech A, Brown LE, Pinto RS. Neuromuscular adaptations to unilateral vs. bilateral strength training in women. J Strength Cond Res 2016;30(7):1924-32. http://dx.doi.org/10.1519/JSC.0000000000001125. PMid:26348920.
3. Lee JH, Cheon S, Jun HP, Huang YL, Chang E. Bilateral comparisons of quadriceps thickness after anterior cruciate ligament reconstruction. Medicina. 2020;56(7):335. http://dx.doi.org/10.3390/medicina56070335. PMid:32635259.
4. Yang JH, Eun SP, Park DH, Kwak HB, Chang E. The effects of anterior cruciate ligament reconstruction on individual quadriceps muscle thickness and circulating biomarkers. Int J Environ Res Public Health 2019;16(24):4895. http://dx.doi.org/10.3390/ijerph16244895. PMid:31817239.
5. O'Brien WD Jr. Ultrasound - biophysics mechanisms. Prog Biophys Mol Biol 2007;93(1-3):212-55. http://dx.doi.org/10.1016/j.pbiomolbio.2006.07.010. PMid:16934858.
6. Larivière C, Gagnon D, Oliveira E Jr, Henry SM, Mecheri H, Dumas JP. Reliability of ultrasound measures of the transversus abdominis: effect of task and transducer position. PM&R 2013;5(2):104-13. http://dx.doi.org/10.1016/j.pmrj.2012.11.002. PMid:23313039.
7. Larivière C, Gagnon D, Oliveira E Jr, Henry SM, Mecheri H, Dumas JP. Ultrasound measures of the lumbar multifidus: effect of task and transducer position on reliability. PM&R 2013;5(8):678-87. http://dx.doi.org/10.1016/j.pmrj.2013.03.010. PMid:23507347.
8. Blazevich AJ, Gill ND, Zhou S. Intra- and intermuscular variation in human quadriceps femoris architecture assessed in vivo. J Anat 2006;209(3):289-310. http://dx.doi.org/10.1111/j.1469-7580.2006.00619.x. PMid:16928199.
9. Enes A, Alves RC, Schoenfeld BJ, Oneda G, Perin SC, Trindade TB, et al. Rest-pause and drop-set training elicit similar strength and hypertrophy adaptations compared with traditional sets in resistance-trained males. Appl Physiol Nutr Metab 2021;46(11):1417-24. http://dx.doi.org/10.1139/apnm-2021-0278. PMid:34260860.
10. Abe T, Dehoyos DV, Pollock ML, Garzarella L. Time course for strength and muscle thickness changes following upper and lower body resistance training in men and women. Eur J Appl Physiol 2000;81(3):174-80. http://dx.doi.org/10.1007/s004210050027. PMid:10638374.
11. Ogasawara R, Thiebaud RS, Loenneke JP, Loftin M, Abe T. Time course for arm and chest muscle thickness changes following bench press training. Interv Med Appl Sci 2012;4(4):217-20. http://dx.doi.org/10.1556/imas.4.2012.4.7. PMid:24265879.
12. Coratella G, Beato M, Bertinato L, Milanese C, Venturelli M, Schena F. Including the eccentric phase in resistance training to counteract the effects of detraining in women: a randomized controlled trial. J Strength Cond Res 2022;36(11):3023-31. http://dx.doi.org/10.1519/JSC.0000000000004039. PMid:34537804.
13. Stock MS, Olinghouse KD, Drusch AS, Mota JA, Hernandez JM, Akalonu CC, et al. Evidence of muscular adaptations within four weeks of barbell training in women. Hum Mov Sci 2016;45:7-22. http://dx.doi.org/10.1016/j.humov.2015.11.004. PMid:26583966.
14. Ema R, Wakahara T, Miyamoto N, Kanehisa H, Kawakami Y. Inhomogeneous architectural changes of the quadriceps femoris induced by resistance training. Eur J Appl Physiol 2013;113(11):2691-703. http://dx.doi.org/10.1007/s00421-013-2700-1. PMid:23949789.
15. Trindade TB, Prestes J, Oliveira L No, Medeiros RMV, Tibana RA, Sousa NMF, et al. Effects of pre-exhaustion versus traditional resistance training on training volume, maximal strength, and quadriceps hypertrophy. Front Physiol 2019;10:1424. http://dx.doi.org/10.3389/fphys.2019.01424. PMid:31824336.
16. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. http://dx.doi.org/10.1136/bmj.n71. PMid:33782057.
17. Mokkink LB, Boers M, van der Vleuten CPM, Bouter LM, Alonso J, Patrick DL, et al. COSMIN risk of bias tool to assess the quality of studies on reliability or measurement error of outcome measurement instruments: a Delphi study. BMC Med Res Methodol 2020;20(1):293. http://dx.doi.org/10.1186/s12874-020-01179-5. PMid:33267819.

18. Betz TM, Wehrstein M, Preisner F, Bendszus M, Friedmann-Bette B. Reliability and validity of a standardized ultrasound examination protocol to quantify vastus lateralis muscle. J Rehabil Med 2021;53(7):jrm00212. http://dx.doi.org/10.2340/16501977-2854. PMid:34121129.

19. Mechelli F, Arendt-Nielsen L, Stokes M, Agyapong-Badu S. Validity of ultrasound imaging versus magnetic resonance imaging for measuring anterior thigh muscle, subcutaneous fat, and fascia thickness. Methods Protoc 2019;2(3):58. http://dx.doi.org/10.3390/mps2030058. PMid:31295936.

20. Nijholt W, Jager-Wittenaar H, Raj IS, van der Schans CP, Hobbelen H. Reliability and validity of ultrasound to estimate muscles: a comparison between different transducers and parameters. Clin Nutr ESPEN 2020;35:146-52. http://dx.doi.org/10.1016/j.clnesp.2019.10.009. PMid:31987109.

21. Worsley PR, Kitsell F, Samuel D, Stokes M. Validity of measuring distal vastus medialis muscle using rehabilitative ultrasound imaging versus magnetic resonance imaging. Man Ther 2014;19(3):259-63. http://dx.doi.org/10.1016/j.math.2014.02.002. PMid:24582328.

22. Barotsis N, Tsiganos P, Kokkalis Z, Panayiotakis G, Panagiotopoulos E. Reliability of muscle thickness measurements in ultrasonography. Int J Rehabil Res 2020;43(2):123-8. http://dx.doi.org/10.1097/MRR.0000000000000390. PMid:31913183.

23. Caresio C, Salvi M, Molinari F, Meiburger KM, Minetto MA. Fully automated muscle ultrasound analysis (MUSA): robust and accurate muscle thickness measurement. Ultrasound Med Biol 2017;43(1):195-205. http://dx.doi.org/10.1016/j.ultrasmedbio.2016.08.032. PMid:27720522.

24. Chiaramonte R, Bonfiglio M, Castorina EG, Antoci SAM. 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 2019;65(2):165-70. http://dx.doi.org/10.1590/1806-9282.65.2.165. PMid:30892439.

25. Takahashi Y, Fujino Y, Miura K, Toida A, Matsuda T, Makita S. Intra- and inter-rater reliability of rectus femoris muscle thickness measured using ultrasonography in healthy individuals. Ultrasound J 2021;13(1):21. http://dx.doi.org/10.1186/s13089-021-00224-8. PMid:33856566.

26. Arruda BR, Soares ALC, Carvalho RF, Gomes PSC. Reliability and measurement error of the proximal, medial and distal portions of the vastus lateralis muscle thickness measured with extended field of view ultrasonography. Kinesiol 2022;54(1):107-14. http://dx.doi.org/10.26582/k.54.1.11.

27. Carr JC, Gerstner GR, Voskuil CC, Harden JE, Dunnick D, Badillo KM, et al. The influence of sonographer experience on skeletal muscle image acquisition and analysis. J Funct Morphol Kinesiol 2021;6(4):91. http://dx.doi.org/10.3390/jfmk6040091. PMid:34842750.

28. Cleary CJ, Nabavizadeh O, Young KL, Herda AA. Skeletal muscle analysis of panoramic ultrasound is reliable across multiple raters. PLoS One 2022;17(5):e0267641. http://dx.doi.org/10.1371/journal.pone.0267641. PMid:35500010.

29. Dudley-Javoroski S, McMullen T, Borgwardt MR, Peranich LM, Shields RK. Reliability and responsiveness of musculoskeletal ultrasound in subjects with and without spinal cord injury. Ultrasound Med Biol 2010;36(10):1594-607. http://dx.doi.org/10.1016/j.ultrasmedbio.2010.07.019. PMid:20800961.

30. Ema R, Wakahara T, Mogi Y, Miyamoto N, Komatsu T, Kanehisa H, et al. In vivo measurement of human rectus femoris architecture by ultrasonography: validity and applicability. Clin Physiol Funct Imaging 2013;33(4):267-73. http://dx.doi.org/10.1111/cpf.12023. PMid:23692615.

31. Franchi MV, Longo S, Mallinson J, Quinlan JI, Taylor T, Greenhaff PL, et al. Muscle thickness correlates to muscle cross-sectional area in the assessment of strength training-

induced hypertrophy. Scand J Med Sci Sports 2018;28(3):846-53. http://dx.doi.org/10.1111/sms.12961. PMid:28805932.

32. Gomes PSC, Meirelles CM, Leite SP, Montenegro CAB. Confiabilidade da medida de espessuras musculares pela ultrassonografia. Rev Bras Med Esporte 2010;16(1):41-5. http://dx.doi.org/10.1590/S1517-86922010000100008.

33. Hagoort I, Hortobágyi T, Vuillerme N, Lamoth CJC, Murgia A. Age- and muscle-specific reliability of muscle architecture measurements assessed by two-dimensional panoramic ultrasound. Biomed Eng Online 2022;21(1):15. http://dx.doi.org/10.1186/s12938-021-00967-4. PMid:35152889.

34. Ishida H, Suehiro T, Suzuki K, Watanabe S. Muscle thickness and echo intensity measurements of the rectus femoris muscle of healthy subjects: intra and interrater reliability of transducer tilt during ultrasound. J Bodyw Mov Ther 2018;22(3):657-60. http://dx.doi.org/10.1016/j.jbmt.2017.12.005. PMid:30100293.

35. Jacob I, Jones G, Francis P, Johnson MI. The effect of limb position on measured values of vastus lateralis muscle morphology using B Mode ultrasound. Transl Sports Med 2021;4(6):697-705. http://dx.doi.org/10.1002/tsm2.271.

36. Lanferdini FJ, Sonda FC, Paz IA, Oliveira LZ, Wagner ES No, Molinari T, et al. Reliability of knee extensor neuromuscular structure and function and functional tests' performance. J Bodyw Mov Ther 2021;27:584-90. http://dx.doi.org/10.1016/j.jbmt.2021.05.004. PMid:34391291.

37. Lima KMM, Oliveira LF. Confiabilidade das medidas de arquitetura do músculo vasto lateral pela ultrassonografia. Mot Rev Educ Fis 2013;19(1):217-23. http://dx.doi.org/10.1590/S1980-65742013000100022.

38. Mairet S, Maïsetti O, Portero P. Homogeneity and reproducibility of in vivo fascicle length and pennation determined by ultrasonography in human vastus lateralis muscle. Sci Sports 2006;21(5):268-72. http://dx.doi.org/10.1016/j.scispo.2006.08.004.

39. Mechelli F, Arendt-Nielsen L, Stokes M, Agyapong-Badu S. Inter-rater and intra-rater reliability of ultrasound imaging for measuring quadriceps muscle and non-contractile tissue thickness of the anterior thigh. Biomed Phys Eng Express. 2019;5(3):037002. http://dx.doi.org/10.1088/2057-1976/ab102f.

40. Oranchuk DJ, Nelson AR, Storey AG, Cronin JB. Variability of regional quadriceps architecture in trained men assessed by B-mode and extended-field-of-view ultrasonography. Int J Sports Physiol Perform 2020;15(3):430-6. http://dx.doi.org/10.1123/ijspp.2019-0050. PMid:31188706.

41. Ruas CV, Pinto RS, Lima CD, Costa PB, Brown LE. Test-retest reliability of muscle thickness, echo-intensity and cross sectional area of quadriceps and hamstrings muscle groups using B-mode ultrasound. Int J Kinesiol Sports Sci 2017;5(1):35-41. http://dx.doi.org/10.7575/aiac.ijkss.v.5n.1p.35.

42. Santos R, Armada-da-Silva PAS. Reproducibility of ultrasound-derived muscle thickness and echo-intensity for the entire quadriceps femoris muscle. Radiography 2017;23(3):E51-61. http://dx.doi.org/10.1016/j.radi.2017.03.011. PMid:28687301.

43. Soares ALC, Nogueira FS, Gomes PSC. Assessment methods of vastus lateralis muscle architecture using panoramic ultrasound: a new approach, test-retest reliability and measurement error. Braz J Kinanthrop Hum Perform 2021;23:e76402. http://dx.doi.org/10.1590/1980-0037.2021v23e76402.

44. Sahathevan S, Khor BH, Yeong CH, Tan TH, Mohaideen AKM, Ng HM, et al. Validity of ultrasound imaging in measuring quadriceps muscle thickness and cross-sectional area in patients receiving maintenance hemodialysis. JPEN J Parenter Enteral Nutr 2021;45(2):422-6. http://dx.doi.org/10.1002/jpen.1867. PMid:32384179.

45. Dupont AC, Sauerbrei EE, Fenton PV, Shragge PC, Loeb GE, Richmond FJR. Real-time sonography to estimate muscle thickness: comparison with MRI and CT. J Clin Ultrasound 2001;29(4):230-6. http://dx.doi.org/10.1002/jcu.1025. PMid:11323778.

46. Juul-Kristensen B, Bojsen-Møller F, Holst E, Ekdahl C. Comparison of muscle sizes and moment arms of two rotator cuff muscles measured by ultrasonography and magnetic resonance imaging. Eur J Ultrasound 2000;11(3):161-73. http://dx.doi.org/10.1016/S0929-8266(00)00084-7. PMid:10874191.

47. Kellis E, Galanis N, Natsis K, Kapetanos G. Validity of architectural properties of the hamstring muscles: correlation of ultrasound findings with cadaveric dissection. J Biomech 2009;42(15):2549-54. http://dx.doi.org/10.1016/j.jbiomech.2009.07.011. PMid:19646698.

48. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Med 1998;26(4):217-38. http://dx.doi.org/10.2165/00007256-199826040-00002. PMid:9820922.

49. Hopkins WG. Measures of reliability in sports medicine and science. Sports Med 2000;30(1):1-15. http://dx.doi.org/10.2165/00007256-200030010-00001. PMid:10907753.