European Journal of Clinical Nutrition (2022) 76:401–409
https://doi.org/10.1038/s41430-021-00955-w

ARTICLE

Metabolism and Metabolomics

How reliably can ultrasound help determine muscle and adipose tissue thickness in clinical settings? An assessment of intra- and inter-examiner reliability in the USVALID study

Arabella Fischer[1] · Ricarda Hahn[1] · Martin Anwar[1] · Anatol Hertwig[1] · Maximilian Pesta[1] · Isabel Timmermann[1] · Timo Siebenrock[1] · Konstantin Liebau[1] · Jonas Brugger[2] · Martin Posch[2] · Isabella Sulz[2] · Michael Hiesmayr[1,2]

Received: 16 December 2020 / Revised: 15 May 2021 / Accepted: 27 May 2021 / Published online: 5 July 2021
© The Author(s), under exclusive licence to Springer Nature Limited 2021

Abstract

Background/Objectives Ultrasound is used to measure muscle and adipose tissue thickness at the bedside. This study was aimed at determining the intra- and inter-examiner reliability for marking points to measure adipose tissue and muscle thickness and assessing it in terms of the performance and evaluation of the corresponding ultrasound scans.

Subjects/Methods Intra- and inter-examiner reliability was tested in 120 patients. Limb lengths were measured to mark three and two measuring points on both the thighs and upper arms, respectively. Ultrasound scans were performed at each measuring point to evaluate muscle and adipose tissue thickness.

Results Regarding the marking of the measuring points, intra- and inter-examiner reliability were high to very high, with correlation coefficients ranging from 0.74 to 0.96. In the performance and evaluation of adipose tissue thickness, all measuring points showed a high to very high reliability, with correlation coefficients ranging from 0.70 to 0.97. In the performance and evaluation of muscle thickness, the ventral measuring point on the thigh and the anterior measuring point on the upper arm showed the best reliability, with high to very high correlation coefficients ranging from 0.77 to 0.93.

Conclusions In terms of intra- and inter-examiner reliability, the ventral measuring point on the thigh and the anterior measuring point on the upper arm can be strongly recommended for ultrasound measurements of muscle and adipose tissue thickness.

Introduction

Muscle and adipose mass are major determinants of outcomes in acute and chronic conditions such as critical illness [1], chronic heart failure [2], and cancer [3, 4]. Muscle and adipose tissue mass are determined by body composition measurements. Changes in body composition are important factors influencing prognosis and therapeutic guidance in terms of nutrition [5], medication dose [3], and physiotherapy [5]. A higher than 5% or 10% decrease in body weight within or beyond 6 months, respectively, is a major indicator of malnutrition [6]. It is important to know about muscle or fat loss [6]. The reference methods of body composition measurement are based on imaging, such as magnetic resonance imaging (MRI) [7], computed tomography (CT) [8, 9], and dual-energy X-ray absorptiometry (DEXA) [8]. Other non-bedside methods are based on plethysmography using displacement of air or water [10]. Bedside evaluation methods are based on ultrasound [11], anthropometry [12], or bioimpedance measurement [8]. Determinations of adipose tissue thickness by using ultrasound are more precise compared to those performed using skinfold measurements [13]. Muscle thickness can also be determined using ultrasound [11, 14]. Ultrasound involves three steps: marking of measuring points, scan performance, and evaluation of muscle and adipose tissue thickness [11]. Each of these steps affects the reliability of the measurement of muscle and adipose tissue thickness.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41430-021-00955-w.

✉ Arabella Fischer
arabella.fischer@meduniwien.ac.at

[1] Division of Cardiothoracic and Vascular Anaesthesia and Intensive Care Medicine, Medical University of Vienna, Vienna, Austria

[2] Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Vienna, Austria

Reliability is influenced by patient- and operator-related factors: patient's position (use of supports and rotation of limbs), the bodyside (one side might be easier to scan than the other one), marking of measuring points (the use of the same site for each scan repetition), the scan performance (tilting of the probe, minimal compression with the probe, and ensuring a sufficient scanning depth to visualize the bone surface), and evaluation of ultrasound scans (correct identification of the muscle fascia and bone surface) [11, 15]. In previous reliability studies, the identification and quantification of the above-mentioned patient- and operator-related factors responsible for the differences in thickness values between repetitions were challenging [15]. Only a minority of researchers repeated the marking of measuring points. Repeating the marking of the measuring points is crucial when analysing intra- or inter-examiner reliability for thickness measurements using ultrasound [15–22]. However, the reliability for marking measuring points has never been assessed on its own. Therefore, the extent to which the repeat marking of measuring points influences the reliability of the thickness measurements remains unclear. Knowledge of the reliability of the ultrasound method is important for determining the method's precision. Furthermore, knowledge of the precision of the method will allow accurate detection of the changes in muscle and adipose tissue thickness. On the basis of their reliability results, Fivez et al. stated that they would detect a 20% decrease in muscle thickness in adult ICU patients and a 30% decrease in paediatric ICU patients [23]. However, they did not state if they repeated the marking of measuring points. Repeat marking of measuring points is crucial in accounting for all the operator-related factors stated above.

Therefore, the aim of this study was to determine the intra- and inter-examiner reliability for marking points to measure adipose tissue and muscle thickness and assessing the reliability in terms of the performance and evaluation of the corresponding ultrasound scans.

Materials/subjects and methods

Study design and population

One hundred and twenty non-critically ill patients were recruited for the reliability assessment of the ultrasound method in the USVALID study at the Medical University of Vienna from 2017 to 2018 (clinicaltrials.gov identifier: NCT03160222). Patients were included in the USVALID study if they had undergone CT at the level of the third lumbar vertebra for any clinical reason. The ultrasound examination had to be performed within 48 h of the CT scan. Patients were excluded if they were younger than 18 years. The study was approved by the Ethics Committee of the Medical University of Vienna and conducted according to the Declaration of Helsinki. Reporting was done according to the STROBE guidelines [24].

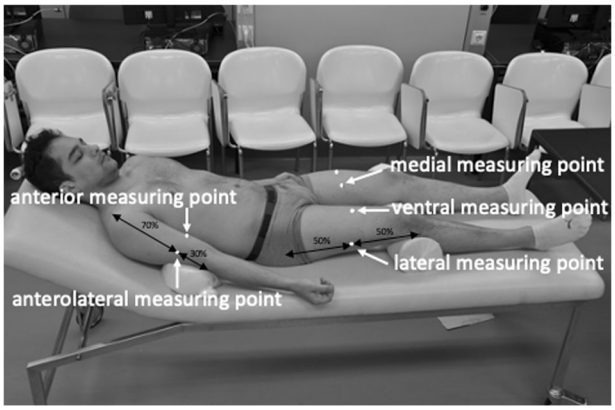

Fig. 1 Marking of measuring points. 3 measuring points are marked on the thigh on both sides. 2 measuring points are marked on the upper arm on both sides. "Reprinted and adapted from Clinical Nutrition Experimental, 32:38–73, Fischer A, Anwar M, Hertwig A, Hahn R, Pesta M, Timmermann I, Siebenrock T, Liebau K, Hiesmayr M, Ultrasound method of the USVALID study to measure subcutaneous adipose tissue and muscle thickness on the thigh and upper arm: an illustrated step-by-step guide, Copyright (2020), with permission from Elsevier" [11].

Ultrasound examination

An illustrated step-by-step guide with all details of the USVALID ultrasound examination has been published [11]. The methodology of the examination is recapitulated in Supplementary Information, Table S1. In brief, the ultrasound examination comprised three steps:

Step 1 – Marking of measuring points: precise anatomical landmarks were palpated to determine the upper arm and thigh length. The anterior and anterolateral measuring points were marked on each side at 70% of the upper arm length. The ventral, lateral, and medial measuring points were marked on each side at 50% of the thigh length (Fig. 1). Furthermore, limb circumferences were measured at 70% of the upper arm length and 50% of the thigh length on each side, and limb lengths and circumferences were noted.

Step 2 – Ultrasound scanning: two scans were performed at each measuring point: one in the short-axis plane and another in the long-axis plane. Scanning in both planes helped identify the muscle fascia, especially when oedema is present [11, 14, 20, 25–27]. Minimal compression was applied using a gel pad and additional gel on top of it (Supplementary Information, Fig. S1). Furthermore, to guarantee minimal compression, it was verified that the borders of each scan were blurred (Supplementary Information, Fig. S1). The Siemens Acuson Freestyle ultrasound scanner was used.

Step 3 – Evaluation of muscle and adipose tissue thickness: muscle and adipose tissue thickness were measured on

each scan by using the inbuilt callipers of the ultrasound machine. The scan quality was evaluated according to the visibility of the muscle fascia and bone surface. In quality 1 scans, the muscle fascia and bone surface were clearly visible. In quality 2 scans, it was possible to spot the muscle fascia and bone surface. In quality 3 scans, the muscle fascia and bone surface were indistinguishable and no evaluation was possible [11]. After the quality assessment, the adipose tissue and muscle thickness were measured at the exact center of the scan by drawing the shortest possible line from the skin to the bone surface. Adipose tissue thickness including that of the skin and muscle fascia was measured. The thickness of the entire muscle without that of the muscle fascia was measured because it is easier to delineate muscle fascia from muscle tissue than to delineate muscle fascia from subcutaneous adipose tissue (Supplementary Information, Fig. S1) [11].

Assessment of intra- and inter-examiner reliability assessment

There were five examiners. Before the study, one experienced examiner trained four novice examiners. The ultrasound examination was performed 20 times in healthy subjects by each examiner. The first and second runs of the ultrasound examination required 2 h in all. To reduce the burden of the examination for the patient, intra- and inter-examiner reliability was assessed in two groups of 60 patients. For assessing intra-examiner reliability, it was necessary that both the first and second runs of the ultrasound examination in 60 patients be performed by the same examiner. Regarding the assessment of inter-examiner reliability, it was necessary that the second run of the ultrasound examination in 60 patients be performed by a different examiner. The first run of the ultrasound examination was performed on both sides. After the first run, all markings of the measuring points were erased with a disinfectant. After the measuring points were erased, the second run was performed by the same or a different examiner on only one randomly chosen body side to reduce the burden of the examination for the patient. During the second run, the examiner (same or different) had to relocate the markings of the measuring points by measuring the limb lengths. The examiner of the second run noted the values of limb length and muscle and adipose tissue thickness on a different case report form to ensure that the examiner was not influenced by the values of the first run.

Statistical analysis

First, intra- and inter-examiner reliability for marking the measuring points was assessed by comparing limb length measurements from the first and second runs. Second, intra- and inter-examiner reliability for the performance and evaluation of scans was assessed by comparing adipose and muscle thickness values from the first and second runs. For all reliability analyses, measures of correlation (correlation coefficient, CC) and magnitude of the error (Bland–Altman plots) were reported [28, 29].

For each measuring point, linear mixed models were computed. These included muscle thickness values from the first and second runs, random effects of the examiner ($n = 5$) and patient ($n = 60$), and a nested effect for the bodyside (right or left). The intra- and inter-examiner CCs were defined as the proportion of variance explained by the patient and body side in the linear mixed model. In mathematical terms, the inter-examiner CC at a certain measuring point was defined as follows:

$$Interexaminer\ CC(muscle\ thickness) = \frac{V(patient)+V(body\ side)}{V(patient)+V(body\ side)+V(examiner)+V(residual)} \quad (1)$$

where V (patient) = estimated variance of muscle thickness values between patients ($n = 60$ patients)

V (body side) = estimated variance of muscle thickness values between the body sides of patients (right or left body side)

V (examiner) = estimated variance of muscle thickness values due to the examiner ($n = 5$ examiners)

V (residual) = residual variance of muscle thickness values

For the assessment of intra-examiner reliability, the same examiner was required to perform both the first and second runs of the ultrasound examination. Thus, the equation for intra-examiner CC only accounts for different patients and body sides and not for different examiners:

$$Intraexaminer\ CC(muscle\ thickness) = \frac{V(patient)+V(body\ side)}{V(patient)+V(body\ side)+V(residual)} \quad (2)$$

where the parameters are defined as above.

The closer the CC was to 1, the more the variance between the first and second ultrasound runs because of the patient rather than the examiner. In other words, the closer the CC was to 1, the better was the intra- or inter-examiner reliability.

Intra- or inter-examiner CCs for adipose tissue thickness at each measuring point for limb length and limb circumference were calculated analogously.

Furthermore, Bland–Altman plots were used to illustrate differences in limb length and muscle and adipose tissue thickness values in scans repeated by the same or different examiner. On the Bland–Altman plots, the 95% lower and upper limits of agreement (LOAs) were respectively calculated as mean difference $-1.96 \times SD$ and mean difference $+1.96 \times SD$. The Bland–Altman

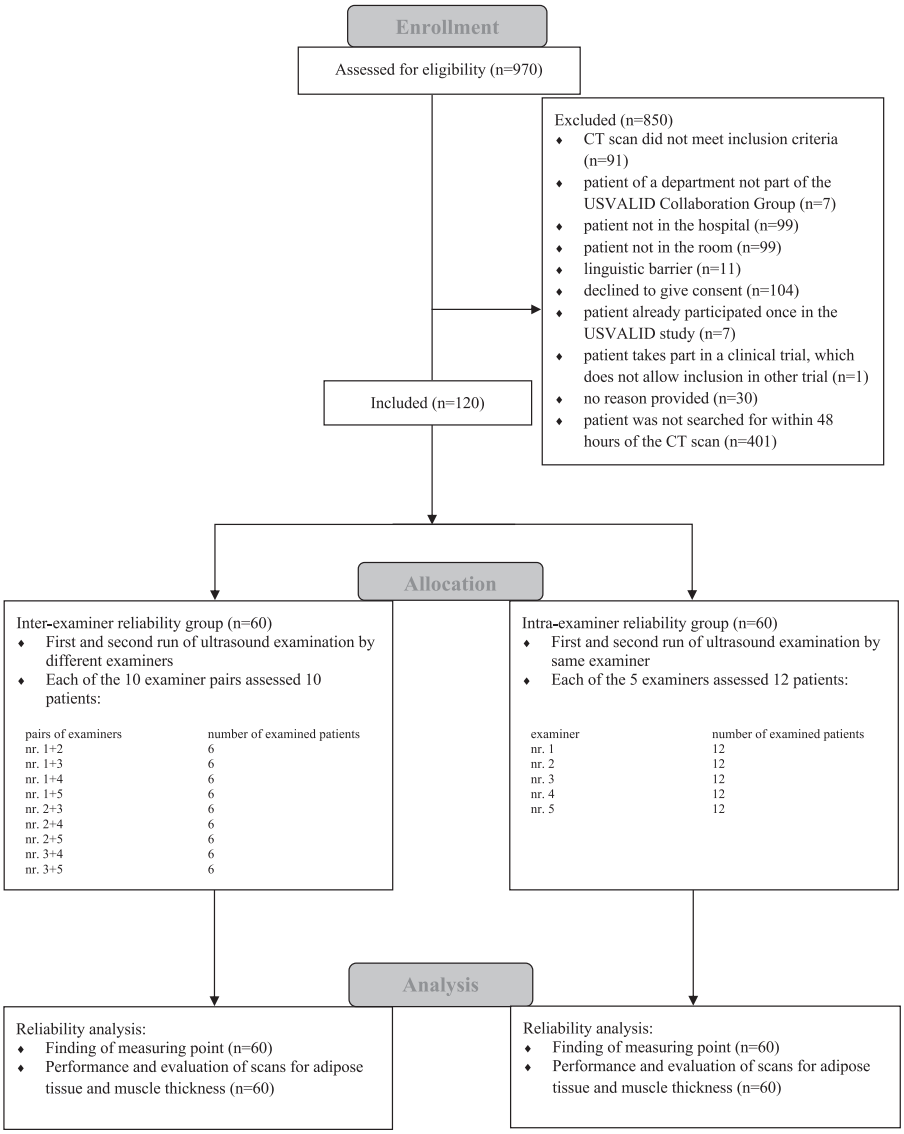

Fig. 2 CONSORT flow diagram for the intra- and inter-examiner reliability assessment of the USVALID study.

plots for intra-examiner reliability display all examiners and those for inter-examiner reliability display all combinations of examiners. R version 3.6.3 (or higher) was used for statistical analysis.

Results

Description of the study population and scans

The CONSORT flow diagram of included patients is presented in Fig. 2. Patients' baseline characteristics are presented in Table 1. A total of 3552 (98.7%) of the 3600 planned scans could be performed in 120 patients, of which 3118 (86.6%) scans had the best quality (quality 1) (Supplementary Information, Table S2). The measuring points that most frequently had the best quality (quality 1) were the anterior measuring point of the upper arm and the ventral measuring point of the thigh. At both measuring points, more than 95% of the scans were quality 1 (Supplementary Information, Fig. S2).

Intra- and inter-examiner reliability for measuring limb lengths, i.e. for marking the measuring points

Intra- and inter-examiner reliability was high to very high for measurements of the upper arm length (intra-examiner CC = 0.96, inter-examiner CC = 0.74) and the thigh length (inter-examiner CC = 0.85, intra-examiner CC = 0.96) (Table 2). For intra-examiner reliability of thigh length measurements, the 95% lower and upper LOA were −1.67 and 1.75 cm, and the corresponding values for inter-examiner reliability were −4.12 and 4.44 cm (Table 2 and Supplementary Information, Fig. S3).

Intra- and inter-examiner reliability for performance and evaluation of scans for adipose tissue and muscle thickness

Intra- and inter-examiner reliability for performance and evaluation of scans depended on the measuring point: For adipose tissue thickness, all measuring points showed high to very high intra- and inter-examiner correlation coefficients ranging from 0.70 to 0.97 (Table 3). For muscle thickness, the lateral and medial measuring points of the thigh, as well as the anterolateral point of the upper arm showed inter-examiner correlation coefficients below 0.65 (Table 4). In contrast, the ventral measuring point of the thigh and the anterior measuring point of the upper arm showed high to very high inter-examiner correlation coefficients ranging from 0.77 to 0.93 (Table 4). For intra-examiner reliability, the 95% lower and upper LOA values were between −0.45 and 0.57 cm compared to −0.81 and 0.80 cm for inter-examiner reliability when performing and evaluating scans for muscle thickness at the ventral measuring point of the thigh (Table 4). Bland-Altman plots for intra- and inter-examiner reliability are presented in the Supplementary Information, Figs. S4-S7.

Table 1 Baseline characteristics of the study population ($n = 120$).

Characteristic	Study population ($n = 120$)
Female	40 (33)
Age (years)	57 ± 17
Weight (kg)	75 ± 15
Height (cm)	173 ± 10
BMI (kg.m^{-2})	25 ± 5
Functional comorbidity index (FCI) (points)	2 (1–4)
Surgical wards	88 (73.3)
General surgery	49 (40.8)
Urology	27 (22.5)
Vascular surgery	4 (3.3)
Cardiac surgery	4 (3.3)
Gynecology	2 (1.7)
Thoracic surgery	1 (0.8)
Orthopedic surgery	1 (0.8)
Medical wards	32 (26.7)
Gastroenterology	22 (18.3)
Cardiology	5 (4.2)
Oncology	4 (3.3)
Hematology	1 (0.8)

Data are indicated as *n* (%) or mean ± SD. FCI is indicated as median (IQR).

Discussion

The intra- and inter-examiner reliability for marking the measuring points was high to very high. The intra- and inter-examiner reliability for performance and evaluation of thickness depended on the measuring points: for adipose tissue thickness, all measuring points showed high to very high intra- and inter-examiner correlations. For muscle thickness, only the ventral measuring point of the thigh and the anterior measuring point of the upper arm showed high to very high intra- and inter-examiner correlations.

Reliability for marking the measuring points

Measurement of the limb length for locating measuring points is a crucial step in ultrasound examinations. Since the intra- and inter-examiner correlations coefficients for measuring limb lengths ranged from 0.74 to 0.96, reliability for measuring limb length was considered to be high to very high. Nevertheless, the 95% upper LOA for intra- and inter-examiner differences between limb length measurements were up to 1.7 cm and 4.4 cm, respectively. Thus, it is not easy to measure limb lengths, even though the anatomical landmarks were defined very precisely to identify the measuring points [11] and examiners were trained before

Table 2 Intra- and inter-examiner reliability for length and circumference measurement on the thigh and upper arm.

			Inter-examiner reliability (in 60 patients)				Intra-examiner reliability (in 60 other patients)			
	Mean (cm)	SD (cm)	CC	Bias (cm)	95% lower LOA (cm)	95% upper LOA (cm)	CC	Bias (cm)	95% lower LOA (cm)	95% upper LOA (cm)
Thigh length	44.8	3.2	0.85	0.16	−4.12	4.44	0.96	0.04	−1.67	1.75
Upper arm length	33.2	2.3	0.74	−0.18	−3.90	3.53	0.96	−0.06	−1.28	1.17
Thigh circumference	49.8	6.1	0.94	−0.16	−4.36	4.05	0.99	−0.17	−2.10	1.76
Upper arm circumference	27.0	3.5	0.94	0.06	−2.61	2.72	0.98	0.03	−1.14	1.20

Mean and SD of all values of both the first and second ultrasound run were calculated in all 120 patients.

CC correlation coefficient, *LOA* limit of agreement.

Table 3 Intra- and inter-examiner reliability for adipose tissue thickness.

				Inter-examiner reliability (in 60 patients)				Intra-examiner reliability (in 60 other patients)			
		Mean (cm)	SD (cm)	CC	Bias (cm)	95% lower LOA (cm)	95% upper LOA (cm)	CC	Bias (cm)	95% lower LOA (cm)	95% upper LOA (cm)
Thigh	Ventral s-a	1.27	0.63	0.94	0.02	−0.44	0.47	0.97	0.02	−0.30	0.33
	Ventral l-a	1.27	0.62	0.94	0.00	−0.45	0.46	0.96	0.00	−0.37	0.37
	Lateral s-a	1.32	0.78	0.94	0.00	−0.54	0.54	0.95	−0.03	−0.60	0.54
	Lateral l-a	1.35	0.80	0.94	−0.03	−0.59	0.53	0.95	−0.01	−0.53	0.51
	Medial s-a	1.74	0.72	0.83	0.03	−0.93	0.99	0.89	−0.04	−0.66	0.58
	Medial l-a	1.72	0.70	0.82	−0.01	−0.98	0.95	0.92	0.02	−0.49	0.54
Upper arm	Anterior s-a	0.65	0.27	0.72	0.03	−0.60	0.65	0.93	0.01	−0.16	0.17
	Anterior l-a	0.66	0.27	0.70	0.00	−0.62	0.63	0.88	0.01	−0.20	0.21
	Anterolateral s-a	1.14	0.54	0.82	−0.01	−0.77	0.74	0.94	0.01	−0.30	0.31
	Anterolateral l-a	1.16	0.54	0.78	−0.01	−0.90	0.89	0.94	0.01	−0.31	0.33

Mean and SD of all values of both the first and second ultrasound run were calculated in all 120 patients.

CC correlation coefficient, *LOA* limit of agreement, *s-a* short-axis plane, *l-a* long-axis plane.

Table 4 Intra- and inter-examiner reliability for muscle thickness.

				Inter-examiner reliability (in 60 patients)				Intra-examiner reliability (in 60 other patients)			
		Mean (cm)	SD (cm)	CC	Bias (cm)	95% lower LOA (cm)	95% upper LOA (cm)	CC	Bias (cm)	95% lower LOA (cm)	95% upper LOA (cm)
Thigh	Ventral s-a	2.50	0.73	0.87	−0.01	−0.81	0.80	0.93	0.06	−0.44	0.57
	Ventral l-a	2.47	0.73	0.90	0.03	−0.69	0.75	0.93	0.06	−0.45	0.57
	Lateral s-a	3.48	0.70	0.59	−0.07	−1.45	1.31	0.74	0.07	−1.08	1.23
	Lateral l-a	3.50	0.73	0.58	0.02	−1.39	1.44	0.77	−0.01	−1.13	1.11
	Medial s-a	2.87	0.76	0.61	−0.15	−1.55	1.26	0.86	0.12	−0.71	0.94
	Medial l-a	2.94	0.78	0.56	0.01	−1.70	1.73	0.79	0.03	−1.01	1.07
Upper arm	Anterior s-a	2.63	0.51	0.77	0.02	−0.74	0.78	0.91	0.09	−0.31	0.50
	Anterior l-a	2.59	0.53	0.77	0.06	−0.70	0.81	0.88	0.01	−0.52	0.54
	Anterolateral s-a	1.54	0.37	0.62	−0.03	−0.77	0.72	0.84	0.05	−0.40	0.50
	Anterolateral l-a	1.49	0.38	0.65	−0.03	−0.75	0.70	0.88	0.04	−0.37	0.44

Mean and SD of all values of both the first and second ultrasound run were calculated in all 120 patients.

CC correlation coefficient, *LOA* limit of agreement, *s-a* short-axis plane, *l-a* long-axis plane.

the study. The development of a sufficient simple method with adequately high reliability is indeed a challenge. Some researchers determined the measuring points from fixed surfaces (wall, table, floor, wooden box) against which the person stands, instead of using bony landmarks [16, 17, 30]. Moreover, they defined the measured distances relative to the person's body height [16, 17, 30]. However, defining distances relative to the person's body height may not always reveal the differing proportions between the upper and lower body in all subjects. Moreover, a standing position is not feasible in an inpatient and intensive care setting. Perin's group used transparent films with individual reference cutaneous marks (beauty spots, scars, and veins) to determine the measuring points again [17]. This is indeed an easy approach that guarantees the same location for multiple measurements over time by different examiners.

Reliability for performance and evaluation of scans

Performance and evaluation of ultrasound scans are other examination steps affecting reliability. For adipose tissue thickness, all measuring points showed high to very high intra- and inter-examiner correlations coefficients ranging from 0.70 to 0.97. For muscle thickness, only the ventral measuring point of the thigh and the anterior measuring point of the upper arm showed high to very high inter-examiner correlation coefficients ranging from 0.77 to 0.93. The ventral measuring point of the thigh and the anterior

measuring point of the upper arm were also the points showing the highest proportion of best quality scans. These two measuring points were easiest to scan because of two reasons: First, they were easily accessible. No tilting of the probe was necessary to visualize muscle fascia and bone [11]. Second, these two measuring points showed a muscular morphology, which is easy to recognize. At less reliable measuring points, the muscular morphology is more complex and variable among patients. Depending on the patient, either the brachialis muscle alone or both the brachialis and the triceps muscle may be visualized at the less reliable anterolateral measuring point of the upper arm [11]. Similarly, the sartorius muscle may or may not be visualized in addition to the quadriceps muscle at the less reliable medial measuring point of the thigh [11]. In summary, the marking of measuring points as the first step of the examination was reliable. Performance and evaluation of the scans as the second and third steps of the examination were reliable at the ventral measuring point of the thigh and the anterior measuring point of the upper arm. Even though there was some variation in the anatomic location of the measuring point, the final thickness values were reliable at these two measuring points.

The reliability for thickness measurements is worse when each examiner independently marks the location of the measuring point instead of performing scans on an already marked measuring point [18]. Therefore, we only compared our results to similar reliability studies, where examiners repeated the marking of measuring points before performing a second scan. The 95% lower and upper LOA found in our study were similar to the LOA in other studies. English et al. noted that the mean muscle thickness at the anterior thigh was 3.1 cm with 95% lower and upper LOAs of −0.88 and 0.72 cm for intra-examiner reliability in stroke patients [15]. We obtained corresponding LOA values of −0.4 to 0.5 cm at the ventral measuring point of the thigh in our study. Paris measured a mean muscle thickness of 3.5 cm over four measuring points on the thigh with 95% lower and upper LOAs of −0.41 and 0.31 cm for inter-examiner reliability in 16 healthy participants [19]. We found corresponding LOA values of −0.7 to 0.8 cm at the ventral measuring point of the thigh in our study. Müller's group measured a sum of eight subcutaneous adipose tissue thicknesses of 0.6 to 7 cm with 95% lower and upper LOAs of ±0.1 to 0.3 cm for intra-examiner reliability in athletes [22]. We found corresponding LOA values of −0.1 to 0.3 cm at the ventral measuring point of the thigh and the anterior measuring point of the upper arm in our study. Paris and Müller calculated the average or the sum of thickness values at different measuring points [16, 19, 22, 31]. A high difference in single thickness values obtained at one difficult measuring point may be less noticeable when averaging all thickness values across different measuring points. Thus, averaging or summing up of findings across multiple measuring points may result in overestimation of the calculated reliability.

Müller et al. excluded skin and muscle fascia when measuring subcutaneous adipose tissue thickness in young athletes [16, 32]. This may be a more precise approach. Müller et al. measured subcutaneous adipose tissue with and without embedded fibrous structures [16, 32]. Reliability was slightly better when measuring subcutaneous adipose tissue with than without embedded fibrous structures [16]. This was because more limits had to be determined to exclude the embedded fibrous structures [16], which is more complex. The limit between the skin and subcutaneous tissue or the limit between the subcutaneous tissue and muscle fascia is often hard to visualize in our experience (Fig. S1) [11]. That is why we included the skin and muscle fascia in the measurement of adipose tissue thickness in our USVALID method. We think that a simple, pragmatic, and reliable method is most important, even though we included skin and muscle fascia in the measurement of adipose tissue thickness. It was our goal to establish a pragmatic and reliable method that can be easily employed not only in healthy individuals but also in hospitalized patients.

Precision

Our reliability results account for the factors related to the marking of the measuring point and the performance and evaluation of adipose tissue and muscle thickness. The 95% LOAs can be considered to represent the degree of precision of our ultrasound examination. The 95% upper LOAs for intra- and inter-examiner reliability of 0.57 cm and 0.80 cm at the ventral measuring point of the thigh would correspond to precision values of 23% and 32% degree in relation to the mean muscle thickness of 2.50 cm (0.57/2.5 = 0.23; 0.80/2.5 = 0.32). This would mean that we can only detect changes in muscle thickness of over 23% and 32%, respectively, when considering the 95% LOAs for intra- and inter-examiner reliability.

Limitations and generalisability

The examiners performing the scans could not be blinded to the thickness values during evaluations on the ultrasound machine. Since the thickness values were directly displayed in the middle of the scan on the ultrasound machine (Siemens Acuson Freestyle) [11], the values could not technically be covered during evaluation. However, the examiner noted the values of the second ultrasound run on a different case report form to avoid the influence of the values noted in the first ultrasound run. We included a large sample size of hospitalized patients from different medical and surgical

specialities. Therefore, our reliability data are applicable to a broad patient population.

Conclusions

Intra- and inter-examiner reliability for marking the measuring points was high to very high. Reliability for performance and evaluation of scans for adipose tissue and muscle thickness was the best at the ventral measuring point of the thigh and the anterior measuring point of the upper arm. Therefore, we recommend measuring adipose tissue and muscle thickness at the ventral measuring point of the thigh and the anterior measuring point of the upper arm.

Acknowledgements We thank all the staff members of the departments of the USVALID Collaboration Group for welcoming our research team to recruit the patients. The representative members of the USVALID Collaboration Group are listed in the Supplementary Information. We thank all the patients who participated in the USVALID study. We thank our friends, listed in the Supplementary Information, who volunteered as healthy individuals during the training phase prior to the trial. The manuscript was edited by a scientific English editing service.

Author contributions AF contributed to the conception, design, patient recruitment, data collection, data analysis and writing. RH, AH, MA, MP, IT, TS, and KL contributed to patient recruitment and data collection. JB and IS contributed to data analysis. MP contributed to conception and data analysis. MH contributed to conception, design, supervision, and writing. All authors contributed to reviewing. All authors read and approved the final manuscript.

Compliance with ethical standards

Conflict of interest The authors declare no competing interests.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Weijs PJM, Looijaard WGPM, Dekker IM, Stapel SN, Girbes AR, Oudemans-van Straaten HM. et al. Low skeletal muscle area is a risk factor for mortality in mechanically ventilated critically ill patients. Crit Care. 2014;18:R12.
2. Fülster S, Tacke M, Sandek A, Ebner N, Tschöpe C, Doehner W, et al. Muscle wasting in patients with chronic heart failure: results from the studies investigating co-morbidities aggravating heart failure (SICA-HF). Eur Heart J. 2013;34:512–9.
3. Prado CM, Lieffers JR, McCargar LJ, Reiman T, Sawyer MB, Martin L, et al. Prevalence and clinical implications of sarcopenic obesity in patients with solid tumours of the respiratory and gastrointestinal tracts: a population-based study. Lancet Oncol. 2008;9:629–35.
4. Martin L, Hopkins J, Malietzis G, Jenkins JT, Sawyer MB, Brisebois R, et al. Assessment of computed tomography (ct)-defined muscle and adipose tissue features in relation to short-term outcomes after elective surgery for colorectal cancer: a multicenter approach. Ann Surgical Oncol. 2018;25:2669–80.
5. Wischmeyer PE, Puthucheary Z, San Millán I, Butz D, Grocott MPW. Muscle mass and physical recovery in ICU: innovations for targeting of nutrition and exercise. Curr Opin Crit Care. 2017;23:269–78.
6. Cederholm T, Jensen GL, Correia MITD, Gonzalez MC, Fukushima R, Higashiguchi T, et al. GLIM criteria for the diagnosis of malnutrition—A consensus report from the global clinical nutrition community. Clin Nutr. 2019;38:1–9.
7. Shen W, Punyanitya M, Wang Z, Gallagher D, St-Onge M-P, Albu J, et al. Total body skeletal muscle and adipose tissue volumes: estimation from a single abdominal cross-sectional image. J Appl Physiol. 2004;97:2333–8.
8. Mourtzakis M, Prado CMM, Lieffers JR, Reiman T, McCargar LJ, Baracos VE. A practical and precise approach to quantification of body composition in cancer patients using computed tomography images acquired during routine care. Appl Physiol, Nutr, Metab. 2008;33:997–1006.
9. Prado CMM, Heymsfield SB. Lean tissue imaging: a new era for nutritional assessment and intervention. J Parenter Enter Nutr. 2014;38:940–53.
10. Gibby JT, Njeru DK, Cvetko ST, Heiny EL, Creer AR, Gibby WA. Whole-body computed tomography-based body mass and body fat quantification: a comparison to hydrostatic weighing and air displacement plethysmography. J Computer Assist Tomogr. 2017;41:302–8.
11. Fischer A, Anwar M, Hertwig A, Hahn R, Pesta M, Timmermann I, et al. Ultrasound method of the USVALID study to measure subcutaneous adipose tissue and muscle thickness on the thigh and upper arm: an illustrated step-by-step guide. Clin Nutr Exp. 2020;32:38–73.
12. Saito R, Ohkawa S, Ichinose S, Nishikino M, Ikegaya N, Kumagai H. Validity of mid-arm muscular area measured by anthropometry in nonobese patients with increased muscle atrophy and variation of subcutaneous fat thickness. Eur J Clin Nutr. 2010;64:899–904.
13. Müller W, Horn M, Furhapter-Rieger A, Kainz P, Kropfl JM, Maughan RJ, et al. Body composition in sport: a comparison of a novel ultrasound imaging technique to measure subcutaneous fat tissue compared with skinfold measurement. Br J Sports Med. 2013;47:1028–35.
14. Fischer A, Spiegl M, Altmann K, Winkler A, Salamon A, Themessl-Huber M, et al. Muscle mass, strength and functional outcomes in critically ill patients after cardiothoracic surgery: does neuromuscular electrical stimulation help? The Catastim 2 randomized controlled trial. Crit Care. 2016;20:30.
15. English CK, Thoirs KA, Fisher L, McLennan H, Bernhardt J. Ultrasound is a reliable measure of muscle thickness in acute stroke patients, for some, but not all anatomical sites: a study of the intra-rater reliability of muscle thickness measures in acute stroke patients. Ultrasound Med Biol. 2012;38:368–76.
16. Müller W, Lohman TG, Stewart AD, Maughan RJ, Meyer NL, Sardinha LB, et al. Subcutaneous fat patterning in athletes: selection of appropriate sites and standardisation of a novel ultrasound measurement technique: ad hoc working group on body composition, health and performance, under the auspices of the IOC Medical Commission. Br J Sports Med. 2016;50:45–54.
17. Perin F, Pittet J-C, Schnebert S, Perrier P, Tranquart F, Beau P. Ultrasonic assessment of variations in thickness of subcutaneous fat during the normal menstrual cycle. Eur J Ultrasound. 2000;11:7–14.
18. Sions JM, Velasco TO, Teyhen DS, Hicks GE. Ultrasound imaging: intraexaminer and interexaminer reliability for multifidus muscle thickness assessment in adults aged 60 to 85 years versus younger adults. J Orthop Sports Phys Ther. 2014;44:425–34.
19. Paris MT, Lafleur B, Dubin JA, Mourtzakis M. Development of a bedside viable ultrasound protocol to quantify appendicular lean tissue mass. J Cachexia, Sarcopenia Muscle. 2017;8:713–26.

20. Strasser EM, Draskovits T, Praschak M, Quittan M, Graf A. Association between ultrasound measurements of muscle thickness, pennation angle, echogenicity and skeletal muscle strength in the elderly. Age. 2013;35:2377–88.
21. 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;5:037002.
22. Müller W, Fürhapter-Rieger A, Ahammer H, Lohman TG, Meyer NL, Sardinha LB, et al. Relative body weight and standardised brightness-mode ultrasound measurement of subcutaneous fat in athletes: an international multicentre reliability study, under the auspices of the ioc medical commission. Sports Med. 2020;50:597–614.
23. Fivez T, Hendrickx A, Van Herpe T, Vlasselaers D, Desmet L, Van den Berghe G, et al. An analysis of reliability and accuracy of muscle thickness ultrasonography in critically ill children and adults. J Parenter Enter Nutr. 2016;40:944–9.
24. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. PLOS Med. 2007;4:1623–7.
25. Grimm A, Teschner U, Porzelius C, Ludewig K, Zielske J, Witte OW, et al. Muscle ultrasound for early assessment of critical illness neuromyopathy in severe sepsis. Crit Care. 2013;17:R227.
26. Cartwright MS, Kwayisi G, Griffin LP, Sarwal A, Walker FO, Harris JM, et al. Quantitative neuromuscular ultrasound in the intensive care unit. Muscle Nerve. 2013;47:255–9.
27. Hill JC, Millán IS. Validation of musculoskeletal ultrasound to assess and quantify muscle glycogen content. A novel approach. Physician Sportsmed. 2014;42:45–52.
28. English C, Fisher L, Thoirs K. Reliability of real-time ultrasound for measuring skeletal muscle size in human limbs in vivo: a systematic review. Clin Rehabilitation. 2012;26:934–44.
29. Nijholt W, Scafoglieri A, Jager-Wittenaar H, Hobbelen JSM, van der Schans CP. The reliability and validity of ultrasound to quantify muscles in older adults: a systematic review. Journal of Cachexia. Sarcopenia Muscle. 2017;8:702–12.
30. Perin F, Perrier C, Pittet JC, Beau P, Schnebert S, Perrier P. Assessment of skin improvement treatment efficacy using the photograding of mechanically-accentuated macrorelief of thigh skin. Int J Cosmet Sci. 2000;22:147–56.
31. Störchle P, Müller W, Sengeis M, Ahammer H, Fürhapter-Rieger A, Bachl N, et al. Standardized ultrasound measurement of subcutaneous fat patterning: high reliability and accuracy in groups ranging from lean to obese. Ultrasound Med Biol. 2017;43:427–38.
32. Müller W, Horn M, Fürhapter-Rieger A, Kainz P, Kröpfl JM, Ackland TR, et al. Body composition in sport: Interobserver reliability of a novel ultrasound measure of subcutaneous fat tissue. Br J Sports Med. 2013;47:1036–43.