

● *Original Contribution*

## STANDARDIZED ULTRASOUND MEASUREMENT OF SUBCUTANEOUS FAT PATTERNING: HIGH RELIABILITY AND ACCURACY IN GROUPS RANGING FROM LEAN TO OBESE

PAUL STÖRCHLE,\* WOLFRAM MÜLLER,\* MARIETTA SENGEIS,\* HELMUT AHAMMER,\*  
ALFRED FÜRHPATER-RIEGER,\* NORBERT BACHL,† SONJA LACKNER,‡ SABRINA MÖRKL,§  
and SANDRA HOLASEK‡

\*Institute of Biophysics, Medical University of Graz, Graz, Austria; †Centre of Sports Science, Department of Sports and Physiological Performance, University of Vienna, Vienna, Austria; ‡Institute of Pathophysiology and Immunology, Medical University of Graz, Graz, Austria; and §Department of Psychiatry and Psychotherapeutic Medicine, Medical University of Graz, Graz, Austria

(Received 26 April 2016; revised 8 September 2016; in final form 9 September 2016)

**Abstract**—A recently standardized ultrasound technique for measuring subcutaneous adipose tissue (SAT) was applied to normal-weight, overweight and obese persons. Eight measurement sites were used: upper abdomen, lower abdomen, erector spinae, distal triceps, brachioradialis, lateral thigh, front thigh and medial calf. Fat compression was avoided. Fat patterning in 38 participants (body mass index: 18.6–40.3 kg m<sup>-2</sup>; SAT thickness sums from eight sites: 12–245 mm) was evaluated using a software specifically designed for semi-automatic multiple thickness measurements in SAT (sound speed: 1450 m/s) that also quantifies embedded fibrous structures. With respect to ultrasound intra-observer results, the correlation coefficient  $\rho = 0.999$  ( $p < 0.01$ ), standard error of the estimate = 1.1 mm and 95% of measurements were within  $\pm 2.2$  mm. For the normal-weight subgroup, the median measurement deviation was 0.43 mm (1.1% of mean thickness), and for the obese/overweight subgroup it was 0.89 mm (0.5%). The eight sites used here are suggested to represent inter-individual differences in SAT patterning. High measurement accuracy and reliability can be obtained in all groups, from lean to overweight and obese, provided that measurers are trained appropriately. (E-mail: [wolfram.mueller@medunigraz.at](mailto:wolfram.mueller@medunigraz.at)) © 2016 The Authors. Published by Elsevier Inc. on behalf of World Federation for Ultrasound in Medicine & Biology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

**Key Words:** Body composition, Subcutaneous adipose tissue, Overweight, Obesity, Ultrasound measurement precision.

### INTRODUCTION

Health, physical performance and body composition are closely related. Overweight and obesity cause major health problems in many countries (Hill et al. 2013). According to the World Health Organization (WHO 2015), more than 1.9 billion adults are overweight, and 600 million of them are obese. Various measurement techniques for assessment of body fat have been developed and applied to groups ranging from underweight to overweight and all classes of obesity (Ackland et al. 2012; Heymsfield et al. 2005). Accurate and reliable measurement methods that can also be used in the field

are required to assess body composition status and to monitor effects of interventions. Diagnostic brightness-mode (B-mode) ultrasound (US) can be used to measure subcutaneous adipose tissue (SAT) patterning and changes in it with a sensitivity not reached with any other method (Müller et al. 2016).

Ultrasound has been used to estimate fat thickness since 1965 (Both et al. 1966; Bullen et al. 1965) and has developed into one of the most promising methods for measuring fat layers in the body, particularly for the determination of SAT (Ackland et al. 2012, Müller and Maughan 2013). Diffraction and minimum pulse length limit lateral and axial resolution approximately to the wavelength used. At 18-MHz probe frequency, about 0.1 mm can be reached, but US attenuation (which increases strongly with frequency) limits the investigable SAT depth to a few centimeters. For thick SAT layers, a

Address correspondence to: Wolfram Müller, Institute of Biophysics, Medical University of Graz, Harrachgasse 21/4, 8010 Graz, Austria. E-mail: [wolfram.mueller@medunigraz.at](mailto:wolfram.mueller@medunigraz.at)

frequency of 6 MHz may be necessary. Lower frequency decreases the image resolution (to about 0.3 mm) and, thus, the tissue border detection error, but in thick tissue layers the choice of correct sound speed (and not US image resolution) is the predominant determinant of accuracy (Müller et al. 2016).

Conventional diagnostic US systems use a sound speed of 1540 m/s for calculating the distance from the probe to the boundary between two tissues. In fat, sound speed is substantially lower at 1450 m/s (Herman 2007). This difference cannot be ignored: Errors of about 6% would occur. Another obstacle that had to be overcome is the compressibility of adipose tissue (Ackland et al. 2012; Müller et al. 2013a, 2013b).

A standardized US imaging and tissue thickness evaluation technique has recently been developed to maximize both accuracy and reproducibility and to ease multiple thickness measurements at a set of eight standardized sites (Müller et al. 2016). This technique can be used in the lab and in the field. Compression artifacts are avoided by using a thick layer of US gel between the probe and the skin. The software determines the SAT borders semi-automatically and measures multiple SAT thickness values in a given US image automatically. This measurement approach also enables quantification of embedded fibrous tissues.

To date, this method has been applied in groups of athletes and normal-weight persons (Ackland et al. 2012; Müller et al. 2013a, 2013b, 2016), but not in the overweight or obese. In this study, this new US method is applied to the latter groups for the first time, and intra-observer reliability is analyzed.

## METHODS

### *Anthropometry*

Anthropometric measurements include body mass ( $m$ ), body height ( $h$ ), sitting height ( $s$ ) and the circumferences of waist and hip. The body mass index ( $[BMI] = m/h^2$ ) and mass index  $MI_1 = 0.53 m/(hs)$  were calculated. The MI takes individual sitting height into account (Müller 2009a, 2009b; Müller et al. 2006). Anthropometric measurements were carried out according to the International Society for the Advancement of Kinanthropometry (Stewart et al. 2011).

### *US sites for measurement of SAT patterning*

Figure 1a illustrates the eight site markings for SAT patterning studies. All sites were marked on the right side of the body, in a standing (upper abdomen [UA], lower abdomen [LA], lateral thigh [LT]) or sitting (erector spinae [ES]) position or with the arm (distal triceps [DT], brachioradialis [BR]) or leg (front thigh [FT] and medial calf [MC]) supported, whereas all US measurements

were made with the participants lying in a supine, prone or rotated position. Marking was described in detail recently (Müller et al. 2016). LT was used for the first time for SAT patterning analysis in the study presented here. The site external oblique (EO) was also measured but was not used in the SAT patterning analysis because of the marking and measurement problems that occurred at this site, particularly in obese persons.

### *Participants and observers*

The study was approved by the ethics committee of the Medical University of Graz (20-295 ex08/09). All participants gave their written consent for anonymous use of their data.

A wide range of SAT thicknesses were covered; participants with BMI values between 18.6 and 40.3  $\text{kg m}^{-2}$  were investigated (12 women, 26 men) (Table 1).

The observers are certified International Society for the Advancement of Kinanthropometry anthropometrists, and both had attended a 2-d advanced-level training course on ultrasound fat measurement by the International Association of Sciences in Medicine and Sports. Both observers had previously measured more than 30 persons with this US method.

### *B-Mode ultrasound imaging of SAT*

To avoid fat compression errors, the US probe was placed above a given site without any pressure by using a thick layer of US gel between the probe and the skin. Typically, about 5 mm of US gel should be seen as a dark band in the US image. Conventional US systems were used. Observer I used a Phillips CX50 with the L12-3 linear probe (Phillips, Phillips Healthcare Austria, Europe, Vienna, Austria). Observer II used a GE Logiq-e, with the linear probes 12 L-RS and 9 L-RS, operated at 6 to 13 MHz (General Electric, GE Healthcare Austria, Europe, Vienna, Austria). US measurements at the eight sites were performed in accordance with the standardized US measurement approach (Müller et al. 2016).

### *Semi-automatic thickness measurement*

Ultrasound images were evaluated interactively using FAT software (FAT 3.2, Rotosport, Europe, Stattegg, Austria), which is specifically designed for multiple semi-automatic evaluations of SAT layer thicknesses. Sound speed was set at 1450 m/s for distance determination in SAT (Herman 2007). SAT segments were evaluated using an automatic distance measurement algorithm (Müller et al. 2013a, 2013b, 2016). Depending on the selected region of interest (ROI), typically 50 to 300 individual thicknesses were automatically measured. Tissue segmentation was controlled visually and could be improved, if necessary, by changing the parameters that influence automatic contour detection. The software also

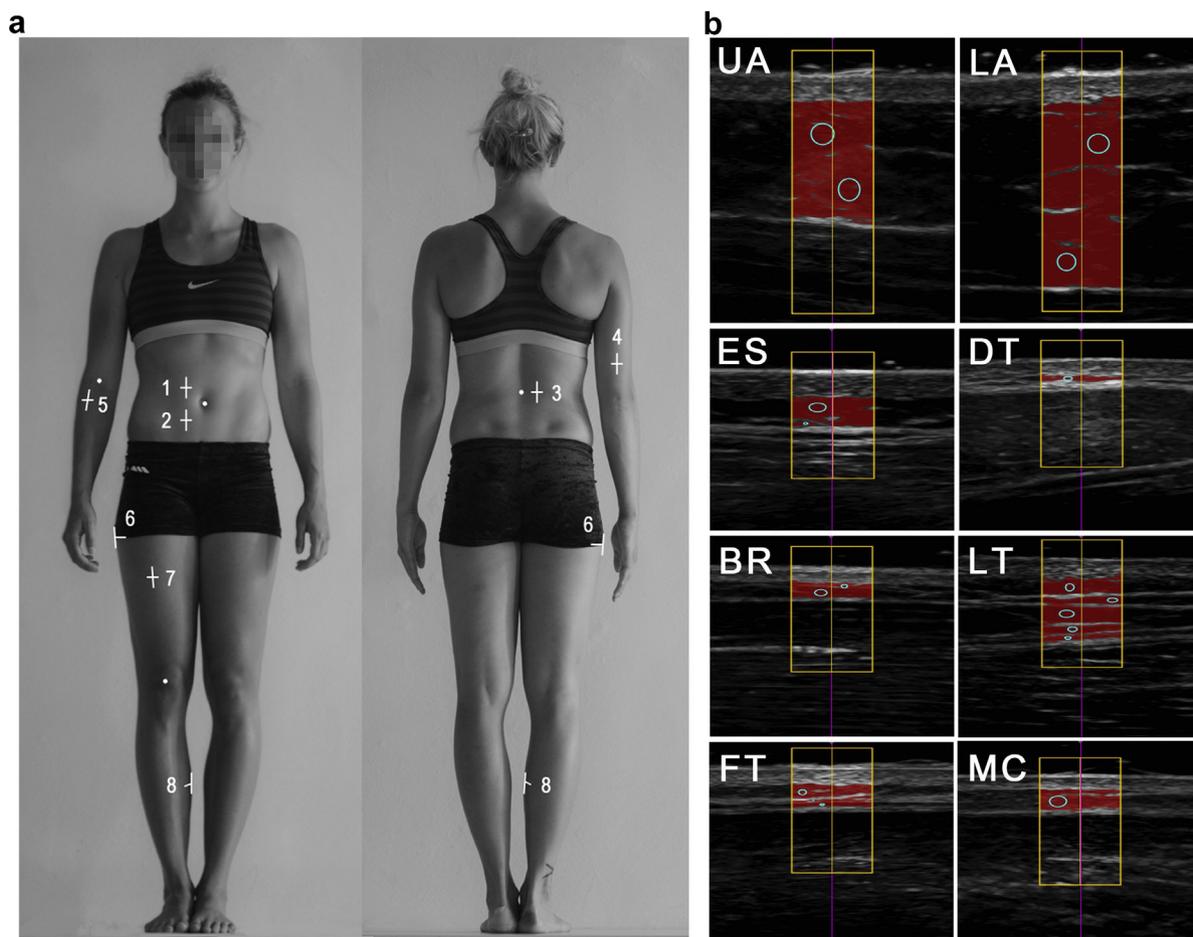


Fig. 1. B-Mode ultrasound measurement of uncompressed subcutaneous fat. (a) Sites for ultrasound measurement of subcutaneous adipose tissue (SAT) patterning: upper abdomen (1, UA), lower abdomen (2, LA), erector spinae (3, ES), distal triceps (4, DT), brachioradialis (5, BR), lateral thigh (6, LT), front thigh (7, FT) and medial calf (8, MC). Body height was used as the reference length for all distances. A detailed description of the standardized marking and of the ultrasound measurement technique can be found in Müller et al. (2016). (b) Ultrasound images and evaluations of SAT thicknesses. Participant A: male, body mass index = 25.5, body mass = 80.7 kg, stature = 1.78 m. (c) Survey plot of SAT patterning according to (b). The columns represent the mean values of the semi-automatic multiple thickness measurements for the eight ultrasound sites. The mean thickness value of the SAT thickness in a given ultrasound image (within the region of interest) is termed  $d_{\text{INCL}}$  (gray) when fibrous structures are included and  $d_{\text{EXCL}}$  (black) when fibrous structures are subtracted. Sum of the eight thicknesses  $D_{\text{INCL}} = 44.3$  mm ( $D_{\text{EXCL}} = 39.8$  mm). (d) Survey plot of a participant B with similar body mass index. Body mass index = 25.4  $\text{kgm}^{-2}$ ; body mass = 75.5 kg, stature = 1.72 m. The SAT thickness sum was 86.9 mm (77.9 mm), 96% higher than in participant A.

enables the operator to distinguish between SAT thicknesses in which fibrous structures are included ( $d_{\text{INCL}}$ ) or excluded ( $d_{\text{EXCL}}$ ), where  $d$  is the SAT thickness at a given site (the average of the distances measured within the ROI). The sums of the eight SAT thicknesses measured in a person are termed  $D_{\text{INCL}}$  and  $D_{\text{EXCL}}$ .

Figure 1b exemplarily illustrates a series of US SAT thickness measurements. The center lines in the US images correspond to the center of the US probe, which was held exactly above the site marking. The rectangular ROI was usually set symmetrically to the center line. In cases in which the thickness of the SAT layer changes when deviating from the center of the site (e.g., at sites UA and at

LT in Fig 1b), a symmetric ROI is necessary to attain high accuracy and reliability. The circles (or ellipses) in the US images indicate the starting zones between skin and muscle fascia where the algorithm searches for the edges (contours) of SAT and measures multiple thicknesses automatically. Finally, a visual control makes sure that the algorithm detected the SAT layer correctly; otherwise, a smaller ROI or a different discrimination parameter for detecting the upper and lower borders of SAT can be chosen.

#### Study protocol

Group 1 (G1, N = 19) consisted of participants who were in the normal BMI range (18.5–25  $\text{kgm}^{-2}$ ) (WHO

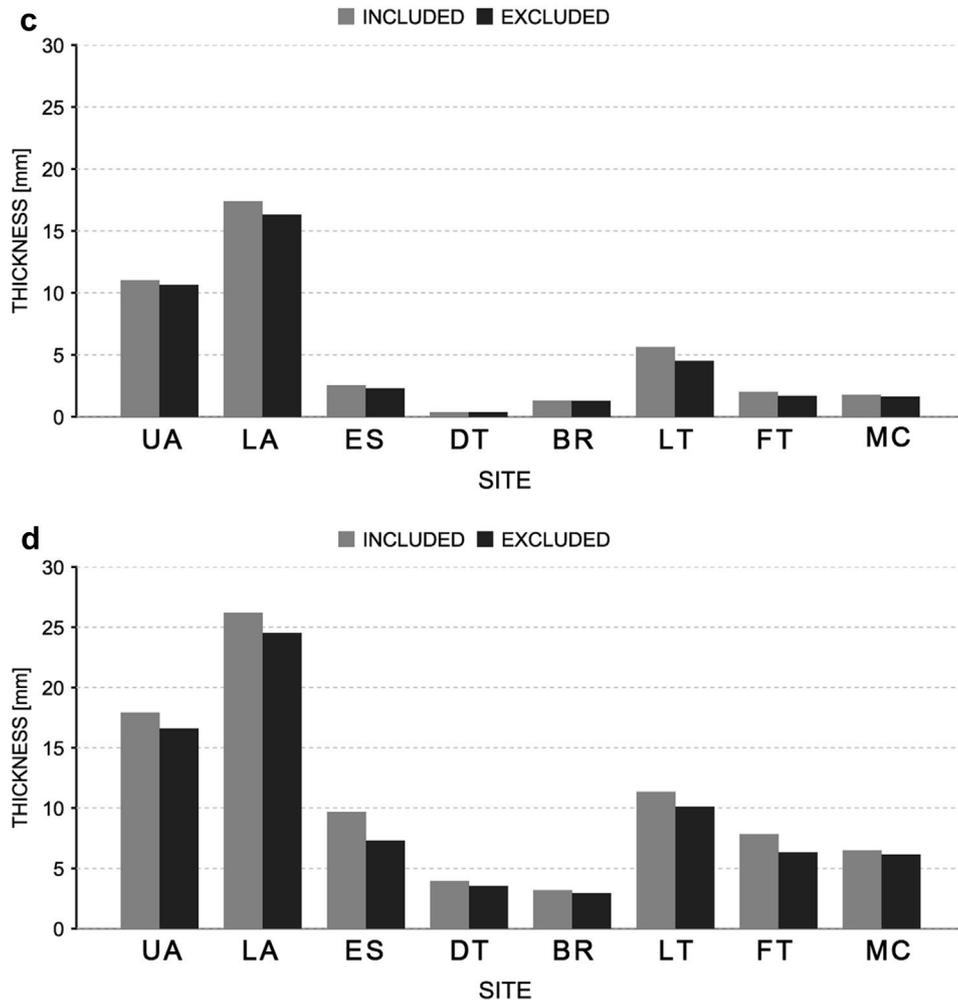


Fig. 1. (continued).

Expert Committee 1995); they were examined by observer I. Group 2 (G2, N = 19, observer II) included the categories overweight (pre-obese 25–30 kgm<sup>-2</sup>, N = 13), obese class I (30–35 kgm<sup>-2</sup>, N = 4), obese class II (35–40 kgm<sup>-2</sup>, N = 1) and obese class III (>40 kgm<sup>-2</sup>, N = 1). The BMI in G2 ranged from 25.4 to 40.3 kgm<sup>-2</sup>. Three measurement series were performed in each participant on 2 separate days within a week. On day 1, the anthropometric and US SAT thickness measurements were conducted (measurement series M1). On the second measurement day, marking was done again before performing M2, and after application of a new layer of gel, the third series M3 followed immediately (using the previous marking of M2).

**Statistics**

SPSS Statistics Version 22 software (IBM, Armonk, NY, USA) was used. Values of the participants are given as the mean ± standard deviation. Shapiro–Wilk tests re-

vealed that not all distributions were normal. Spearman’s rank correlation coefficient (ρ) was used for data in Table 2 and Figure 2. The standard error of estimate (SEE) was used for data in Figure 2. Boxplots were used to sketch the distributions of measurement differences.

Table 1. Characteristics of the participants

	G1		G2		G1 + G2	
	Mean	SD	Mean	SD	Mean	SD
N	19		19		38	
Age (y)	32.5	8.3	47.4	15.6	39.9	14.5
Height (m)	1.75	0.08	1.73	0.08	1.74	0.08
Sitting height <i>s</i> (m)	0.93	0.04	0.91	0.04	0.92	0.04
Mass, <i>m</i> (kg)	68.7	10.4	88.3	14.9	78.5	16.2
Body mass index, BMI (kgm <sup>-2</sup> )	22.2	1.9	29.4	4.2	25.8	4.9
Mass index, MI <sub>I</sub> (kgm <sup>-2</sup> )	22.3	2.1	29.5	4.0	25.9	4.8

SD = standard deviation.

Table 2. Intra-observer correlations\*

US site	Correlation of measurements for $d_{\text{INCL}}^{\dagger}$			Correlation of measurements for $d_{\text{EXCL}}^{\ddagger}$		
	$\rho < 0.01$ (Spearman's rho)			$\rho < 0.01$ (Spearman's rho)		
	$P_{\text{M1,M2}}$	$P_{\text{M1,M3}}$	$P_{\text{M2,M3}}$	$P_{\text{M1,M2}}$	$P_{\text{M1,M3}}$	$P_{\text{M2,M3}}$
UA	0.995	0.994	0.996	0.997	0.997	0.999
LA	0.997	0.996	0.999	0.997	0.996	0.995
ES	0.968	0.971	0.994	0.980	0.973	0.982
DT	0.940	0.930	0.956	0.973	0.939	0.970
BR	0.975	0.975	0.993	0.985	0.988	0.985
LT	0.991	0.988	0.995	0.995	0.991	0.995
FT	0.994	0.989	0.998	0.988	0.987	0.990
MC	0.982	0.982	0.989	0.970	0.965	0.989
Mean	0.980	0.978	0.990	0.986	0.980	0.988

UA = upper abdomen; LA = lower abdomen; ES = erector spinae; DT = distal triceps; BR = brachioradialis; LT = lateral thigh; FT = front thigh; MC = medial calf.

\* Spearman's rank correlation coefficients ( $\rho$ ) were calculated ( $p < 0.01$ ) for each individual site.

<sup>†</sup> Correlations with fibrous structures included. The highest correlations were found at the site lower abdomen (LA), and the lowest at the site distal triceps (DT). All correlations found for the measurements M2 and M3 were higher than the others.

<sup>‡</sup> Correlations with fibrous structures excluded: The highest correlations were found at the UA, and the lowest at the sites DT and MC. In most cases, highest correlations were between the measurements M2 and M3.

## RESULTS

### US images and semi-automatic evaluations of SAT thicknesses at eight standardized sites

At all eight standardized US sites, the upper black layer in the image corresponds to the thick layer of gel between the probe and the skin. Below the gel, epidermis and dermis form a light band. The SAT layer is imaged as a dark band between the lower contour of the skin and the upper contour of the fascia of the muscle that is situated

below the SAT. The US probe is always oriented in a longitudinal direction to the muscle. Fibrous structures (light bands) may be embedded in the SAT layer. All SAT images at the eight standardized sites are of the same structure. The series of evaluated images in Figure 1b (colored areas represent SAT cross-sectional areas within the chosen region of interest) corresponds to the survey plot representing SAT patterning (Fig. 1c). Gray columns represent the thicknesses from the eight sites with the embedded fibrous structures included ( $d_{\text{INCL}}$ ), and black columns represent

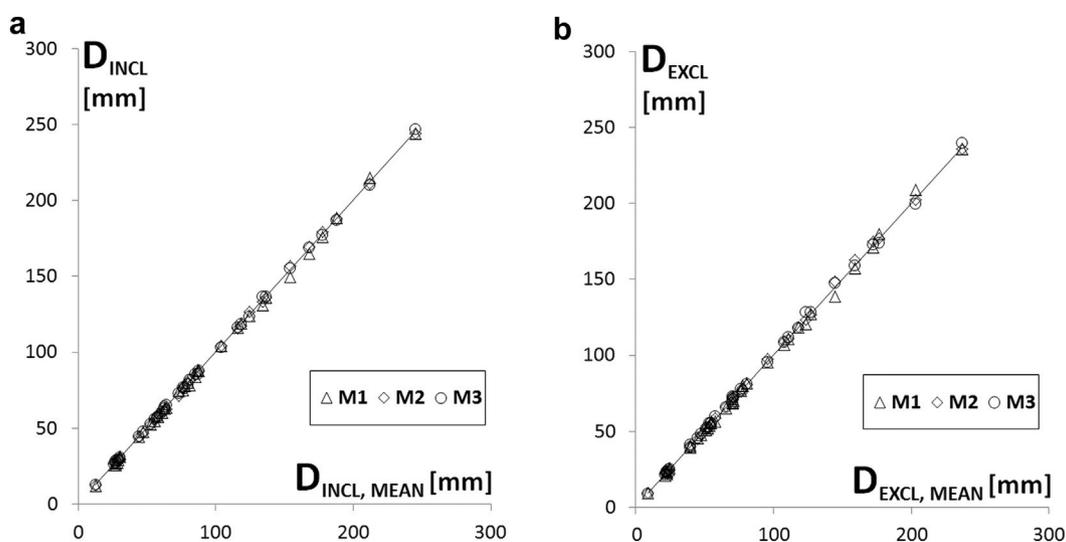


Fig. 2. Sums of subcutaneous adipose tissue (SAT thicknesses) from eight sites measured three times (measurements M1, M2, and M3) in 38 participants. The sums  $D$  of each measurement series at the eight sites (in a given participant) are displayed over the mean value of all three measurements. (a) Embedded fibrous structures included ( $D_{\text{INCL}}$ ). Spearman's rank correlation coefficient ( $\rho$ ) was used to determine the correlation.  $\rho = 0.999$  ( $p < 0.01$ ), standard error of the estimate = 1.1 mm. (b) Embedded fibrous structures excluded ( $D_{\text{EXCL}}$ ).  $\rho = 0.997$  ( $p < 0.01$ ), standard error of the estimate = 1.5 mm.

thicknesses when embedded structures are subtracted ( $d_{\text{EXCL}}$ ). The SAT thickness sum  $D_{\text{INCL}}$  in this participant (BMI = 25.5 kgm<sup>-2</sup>) was 44.3 mm. For comparison, a pattern obtained from another participant with almost the same BMI (25.4 kgm<sup>-2</sup>) is provided in Figure 1d;  $D_{\text{INCL}}$  was 86.9 mm in the latter case. This example indicates that the BMI cannot be used as a measure of body fat: The SAT sums differ by almost a factor of 2 (96% difference).

#### Intra-observer study

Figure 2a illustrates the three sums of  $D_{\text{INCL}}$  obtained from the eight sites for the three measurements (M1, M2 and M3). Values are plotted over the mean value of the three measurement sums (in a given participant). Figure 2b illustrates the results for  $D_{\text{EXCL}}$  (without fibrous structures). Statistical characteristics for  $D_{\text{INCL}}$  are  $\rho = 0.999$  ( $p < 0.01$ ) and  $\text{SEE} = 1.1$  mm, and for  $D_{\text{EXCL}}$ ,  $\rho = 0.997$  ( $p < 0.01$ ) and  $\text{SEE} = 1.5$  mm. Several additional borders have to be determined to measure the additional thicknesses of embedded structures for  $D_{\text{EXCL}}$ , whereas for  $D_{\text{INCL}}$ , deviations were slightly smaller because only one upper border and one lower border must be determined by the algorithm.

The corresponding correlations between the individual sites of the single measurements are summarized in Table 2. Spearman's rank correlation coefficient ( $\rho$ ) was calculated. Including fibrous structures, the highest correlation exists for the site LA. With fibrous structures

excluded, UA had the highest correlation. M2 and M3 had the strongest correlations (measurements M2 and M3 were done without re-marking the site).

The SAT thickness sums  $D_{\text{INCL,MEAN}}$  ranged from 12.46 to 77.41 mm in G1 and from 44.34 to 244.87 mm in G2. In both groups (most participants physically untrained), the thickness sums  $D_{\text{INCL,MEAN}}$  for women (N = 12) ranged from 52.51 to 244.87 mm (mean: 117.3 mm), and those for men (N = 26) ranged from 12.46 to 167.88 mm (mean: 72.8 mm).

Figure 3a and b illustrate the deviations of the observers' individual measurement sums ( $\Delta$ ) from their means for each participant. Ninety-five percent of deviations were within the interval  $-2.2$  mm to  $1.9$  mm with fibrous structures included and within  $-3.24$  mm to  $3.18$  mm with fibrous structures excluded.

Boxplots in Figure 4a and b visualize the absolute values of measurement deviations from their mean. The two observers measured the sums of eight sites three times in their respective groups (G1 and G2, respectively), each with 19 participants (resulting in  $2 \times 19 \times 3 = 114$  comparisons of SAT thickness sums). The deviations in the individual groups and in both groups combined are displayed. In G1, the median of the absolute deviations  $\text{ABS}(\Delta_{\text{INCL}})$  was 0.43 mm, and the median of  $\text{ABS}(\Delta_{\text{EXCL}})$  was 0.41 mm; the interquartile ranges were 0.70 and 0.45 mm, respectively. The highest deviations of the single measurement sums of thicknesses were  $\text{ABS}(\Delta_{\text{INCL,max}}) = 1.63$  mm and  $\text{ABS}(\Delta_{\text{EXCL,max}}) = 2.13$  mm; 95% of data were below 1.44 mm (1.57 mm). In G2, the medians were in both

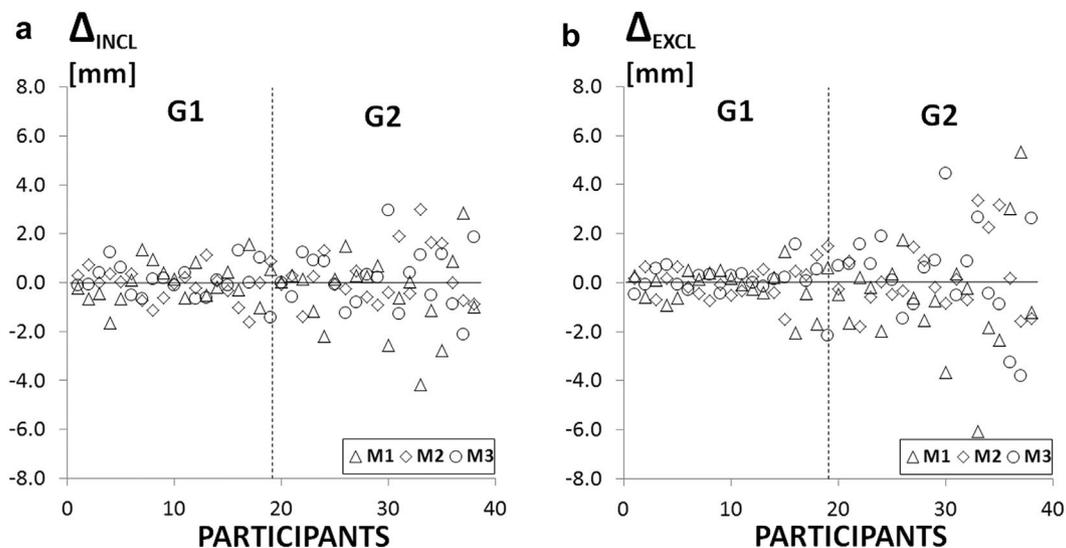


Fig. 3. Differences from the mean of the three measurements (M1, M2 and M3). Three measurement deviations of the sums from the eight sites in each of the 38 participants (both groups combined) are plotted (N = 114). (a)  $\Delta_{\text{INCL}} = D_{\text{INCL}} - D_{\text{INCL,MEAN}}$  is shown for all 38 participants. The 19 participants of each group are ordered according to increasing values of  $D_{\text{INCL,MEAN}}$ . Ninety-five percent of deviations were within the interval  $-2.2$  to  $1.9$  mm. (b)  $\Delta_{\text{EXCL}} = D_{\text{EXCL}} - D_{\text{EXCL,MEAN}}$  is shown for all 38 participants. The 19 participants of each group are ordered according to increasing values of  $D_{\text{INCL,MEAN}}$ . Ninety-five percent of deviations were within the interval  $-3.24$  to  $3.18$  mm.

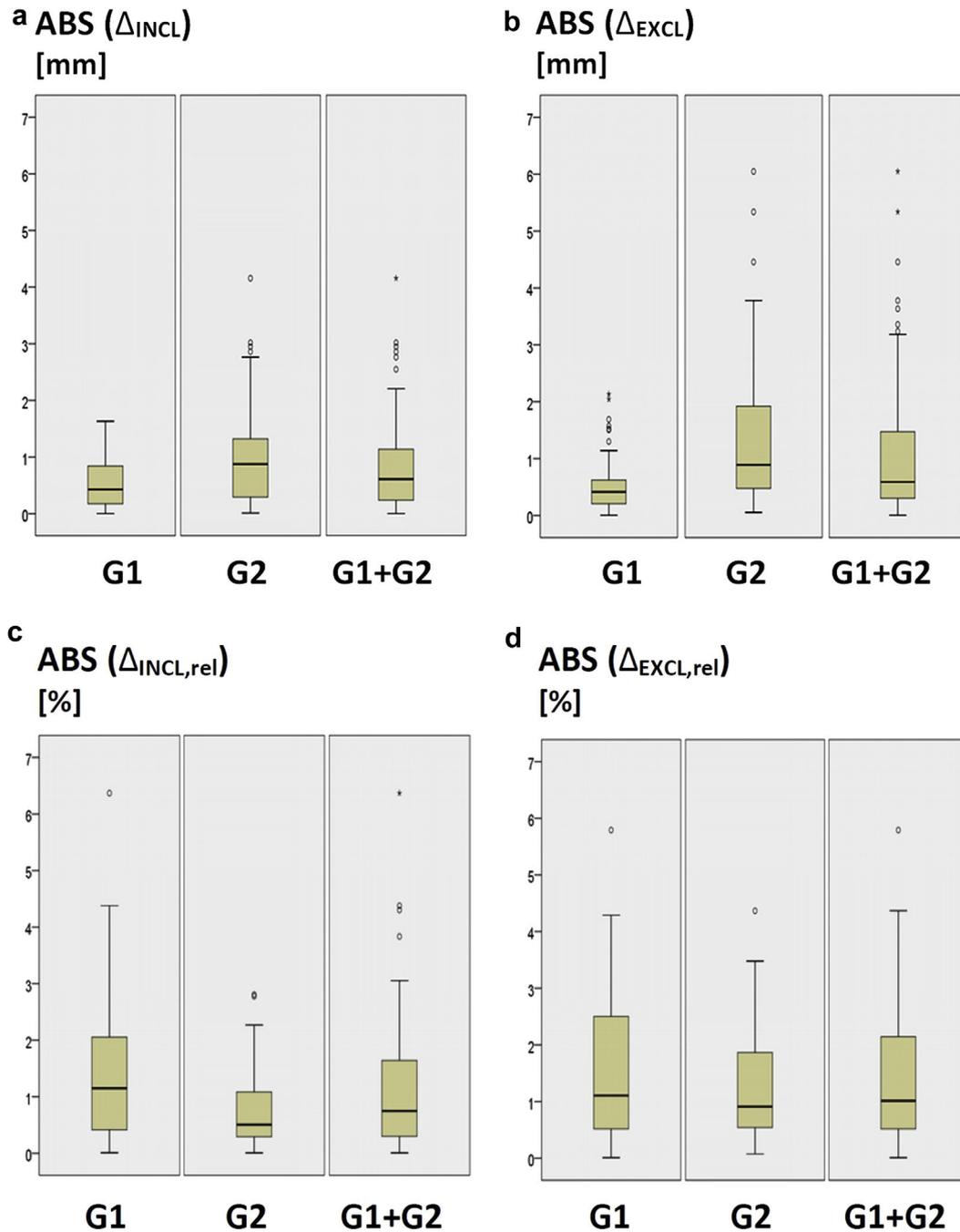


Fig. 4. Observer differences in their individual  $D$  values from their means ( $D_{Mean}$ ). Differences of the three sums from the eight sites in each of the 38 participants (G1:  $N = 57$ , G2:  $N = 57$ , G1 + G2:  $N = 114$ ). (a) Absolute differences:  $ABS(\Delta_{INCL}) = ABS(D_{INCL} - D_{INCL,MEAN})$  and (c) corresponding relative deviations in percent of  $D_{INCL,MEAN}$ :  $\Delta_{INCL,rel} = 100 \cdot ABS(\Delta_{INCL}) / D_{INCL,MEAN}$ . (b) Absolute differences:  $ABS(\Delta_{EXCL}) = ABS(D_{EXCL} - D_{EXCL,MEAN})$  and (d) corresponding relative deviations in percent of  $D_{EXCL,MEAN}$ :  $\Delta_{EXCL,rel} = 100 \cdot ABS(\Delta_{EXCL}) / D_{EXCL,MEAN}$ . Circles represent outliers, and stars represent extreme values.

cases 0.89 mm. The highest deviations were  $ABS(\Delta_{INCL,max}) = 4.16$  mm and  $ABS(\Delta_{EXCL,max}) = 6.05$  mm; the interquartile ranges were 1.06 and 1.48 mm, respectively; 95% of data were below 2.86 mm (3.78 mm).  $ABS(\Delta_{INCL})$  For G1 and G2 together,

median = 0.61 mm and interquartile range = 0.91 mm, and 95% were below 2.21 mm. For  $ABS(\Delta_{EXCL})$ , median = 0.59 mm and interquartile range = 1.18 mm, and 95% were below 3.24 mm. Deviations in the normal-weight group G1 were lower than those in G2.

Figure 4c and d represent the distribution of relative errors  $\Delta_{rel} = 100 \cdot \Delta / D_{MEAN}$  (in percent). For G1, the medians for  $\Delta_{INCL,rel}$  and  $\Delta_{EXCL,rel}$  were both 1.1%. Maximum values were 6.4% and 5.8%. Ninety-five percent were below 3.83% and 3.42%, respectively. For G2, the median of  $\Delta_{INCL,rel}$  was 0.5%, the maximum was 2.8% and 95% of the values were below 2.16%. For  $\Delta_{EXCL,rel}$ , the median was 0.9%, the maximum was 4.4% and 95% were below 3.03%. Deviations in G2 were lower than those in the normal-weight group G1.

For both groups together for  $\Delta_{INCL,rel}$  (and  $\Delta_{EXCL,rel}$ ), median = 0.75% (1.0%), and 95% of the data were below 2.78% (3.42%).

In Figure 5a and b, the absolute values of deviations of the three measurements at each of the eight sites are illustrated. The deviation  $\delta$  is the measurement deviation of each measurement (M1, M2, M3) from the mean of the three measurements at a given site in a given subject. The number of deviations ABS( $\delta$ ) in the boxplots was 114 (38 participants, three measurements). ABS( $\delta_{INCL}$ ) refers to SAT thickness values with the fibrous structures included, and ABS( $\delta_{EXCL}$ ) refers to SAT thickness values without fibrous structures. Table 3 summarizes the statistical characteristics of the boxplots in Figure 5. For both groups combined, median values of ABS( $\delta_{INCL}$ ) ranged from 0.13 to 0.26 mm; for G1, from 0.07 to 0.18 mm; and for G2, from 0.18 to 0.42 mm. Interquartile ranges and maximum deviations can also be found in Table 3.

Figure 6a illustrates median SAT thicknesses and Figure 6b depicts measurement deviations relative to the median SAT thicknesses ( $\delta_{rel}$ ). Although the deviation values ABS( $\delta$ ) increased with increasing  $d$ , the relative values were lower at those sites where SAT thickness  $d$  was higher. The median SAT thickness at the new site LT (which replaces the previously used external oblique site) was between the values for UA and LA (Fig. 6a, gray column). The relative measurement deviations  $\delta_{INCL,rel}$  were below 2%; this is comparable to the values obtained at UA and LA. At all other sites,  $\delta_{INCL,rel}$  was higher (Fig. 6b).

A comparison of BMI and  $D_{INCL}$  is illustrated in Figure 7. There was no correlation in group G1; in G2, there was a moderate correlation of  $\rho = 0.58$  ( $p < 0.01$ ), and for both groups together ( $N = 38$ ),  $\rho = 0.727$  ( $p < 0.01$ ). The highest  $D_{INCL}$  in this group (245 mm, at a BMI of 37.2) was 46% above  $D_{INCL}$  of the person with the highest BMI of 40.3 ( $D_{INCL} = 168$  mm).

DISCUSSION

All SAT images at the eight sites (Fig. 1a) reveal the same structure: skin, SAT, muscle fascia and muscle. To attain the highest accuracy and reliability possible with this US method, it is necessary to mark the eight standardized sites and to apply the US image capturing and

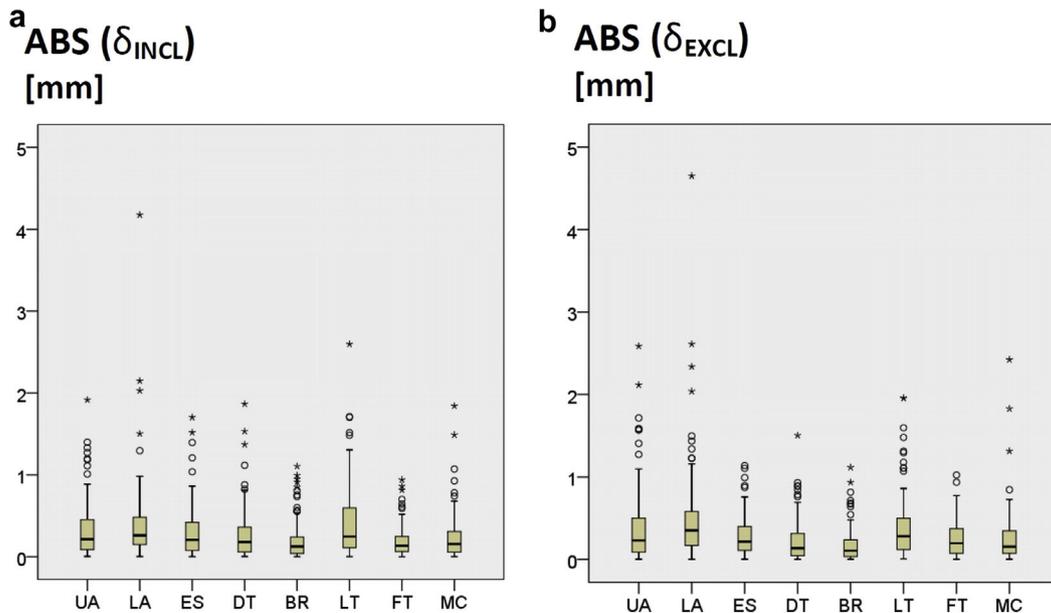


Fig. 5. Measurement differences (absolute values) at the individual eight sites. Absolute differences ABS( $\delta$ ) from both groups together were used. Number of comparisons at each of the eight sites:  $N = 114$  (38 participants, three measurements each). Characteristic values are listed in Table 3. (a) ABS( $\delta_{INCL}$ ) for each of the eight sites. (b) ABS( $\delta_{EXCL}$ ) for each of the eight sites. UA = upper abdomen; LA = lower abdomen; ES = erector spinae; DT = distal triceps; BR brachioradialis; LT = lateral thigh; FT = front thigh; MC = medial calf. Circles represent outliers, and stars represent extreme values.

Table 3. Absolute thickness value differences from the mean of the three measurements for each of the eight sites\*

Measurements	UA	LA	ES	DT	BR	LT	FT	MC
ABS( $\delta_{\text{INCL}}$ ): G1 + G2, N = 3 × 38 = 114 at each site								
Median (mm)	0.21	0.26	0.21	0.18	0.13	0.24	0.13	0.15
Interquartile range (mm)	0.37	0.34	0.35	0.30	0.20	0.49	0.18	0.26
Maximum (mm)	1.9	4.2	1.7	1.9	1.1	2.6	0.9	1.8
ABS( $\delta_{\text{EXCL}}$ ): G1 + G2								
Median (mm)	0.23	0.35	0.22	0.14	0.11	0.28	0.20	0.15
Interquartile range (mm)	0.42	0.41	0.30	0.27	0.20	0.38	0.30	0.27
Maximum (mm)	2.6	4.6	1.1	1.5	1.1	2	1	2.4
ABS( $\delta_{\text{INCL}}$ ): G1, N = 57								
Median (mm)	0.13	0.18	0.11	0.09	0.07	0.16	0.11	0.08
Interquartile range (mm)	0.17	0.21	0.20	0.16	0.13	0.29	0.14	0.15
Maximum (mm)	0.6	0.68	1.4	0.6	0.6	0.85	0.6	0.36
ABS( $\delta_{\text{EXCL}}$ ): G1								
Median (mm)	0.15	0.28	0.14	0.08	0.06	0.26	0.15	0.11
Interquartile range (mm)	0.17	0.28	0.2	0.2	0.11	0.32	0.24	0.18
Maximum (mm)	0.7	0.9	0.87	0.6	0.4	0.86	1	0.57
ABS( $\delta_{\text{INCL}}$ ): G2, N = 57								
Median (mm)	0.42	0.42	0.32	0.31	0.18	0.37	0.18	0.24
Interquartile range (mm)	0.49	0.50	0.50	0.37	0.30	0.72	0.22	0.31
Maximum (mm)	1.9	4.2	1.7	1.9	1.1	2.6	0.9	1.8
ABS( $\delta_{\text{EXCL}}$ ): G2								
Median (mm)	0.43	0.52	0.30	0.27	0.17	0.35	0.27	0.29
Interquartile range (mm)	0.53	0.63	0.42	0.41	0.29	0.57	0.30	0.37
Maximum (mm)	2.6	4.6	1.1	1.5	1.1	2	0.9	2.4

UA = upper abdomen; LA = lower abdomen; ES = erector spinae; DT = distal triceps; BR = brachioradialis; LT = lateral thigh; FT = front thigh; MC = medial calf.

\* Characteristic values of box plots in Figure 5a and b. Deviations of thickness measurements at individual sites with both fibrous structures included and excluded.

evaluation technique as described in the standardization publication by Müller *et al.* (2016). Fibrous structures embedded in the SAT can also be quantified because these structures are clearly visible in the US image and can therefore be easily detected by the contour detection

and evaluation algorithm. Embedded structures were found in many images. The SAT contour detection has to be “semi-automatic” (a visual control and interactive evaluation parameter setting cannot be replaced by a fully automatic algorithm); otherwise border detection errors

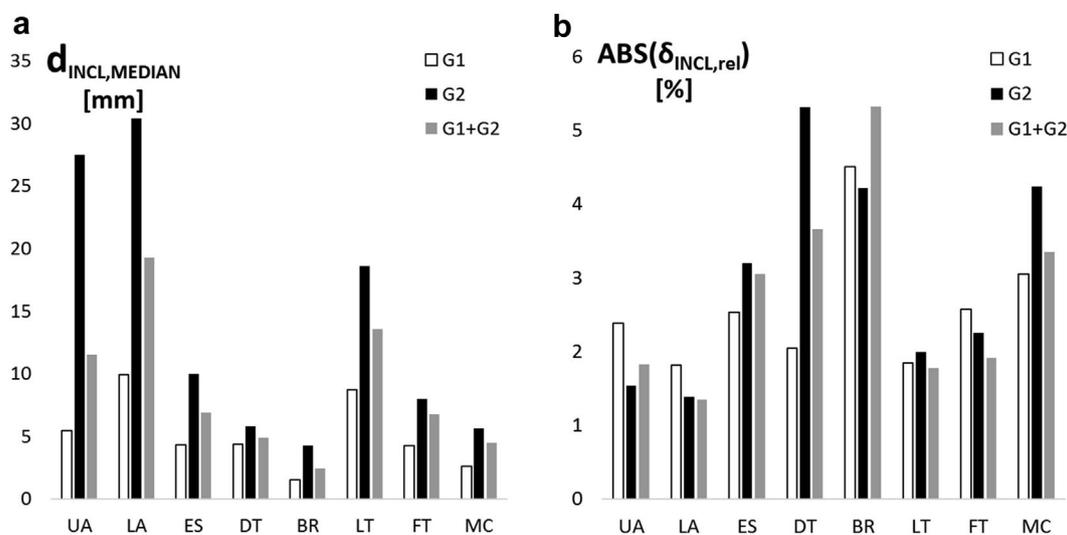


Fig. 6. Subcutaneous adipose tissue (SAT) patterning in the groups and relative measurement deviations at the individual sites: *White columns* correspond to group G1 (N = 57), *black columns* to G2 (N = 57) and *gray columns* to G1 and G2 combined (N = 114). (a) Median SAT thicknesses. (b) Measurement deviations relative to the median SAT thicknesses  $d$ :  $\delta_{\text{INCL,rel}} = 100 \cdot \text{ABS}(\delta_{\text{INCL, MEDIAN}}) / d_{\text{INCL, MEDIAN}}$ , where  $\text{ABS}(\delta_{\text{INCL, MEDIAN}})$  = absolute values of deviations used to determine the median. UA = upper abdomen; LA = lower abdomen; ES = erector spinae; DT = distal triceps; BR = brachioradialis; LT = lateral thigh; FT = front thigh; MC = medial calf.

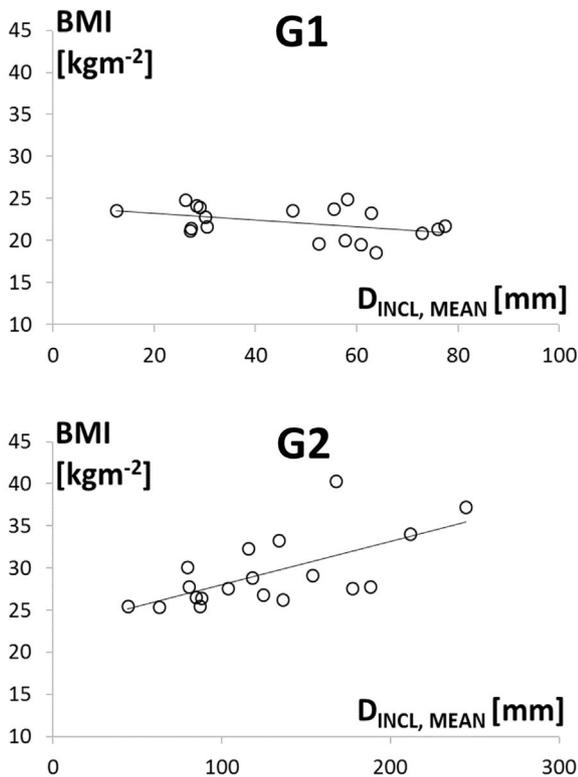


Fig. 7. Comparison of body mass index (BMI) and  $D_{\text{INCL}}$  in groups 1 (G1) and 2 (G2).

would occur because of intermediate structures—for example, Camper’s fascia in the abdominal region (Müller et al. 2013a, 2013b, 2016) or embedded fibrous structures as illustrated in Figure 1b at LT and FT.

Figure 1b exemplarily illustrates the series of evaluated images of participant A, and Figure 1c, the corresponding SAT patterning. Although participant A had almost the same BMI as participant B, the sums of SAT thicknesses differ by almost 100% (Fig. 1d), indicating that valid measures of fat cannot be replaced by simple anthropometric indices such as the BMI (Ackland et al. 2012). This is also evident when studying Figure 7: In group G1, there was no correlation between BMI and the SAT thickness sum  $D$  at all, and also the correlation in G2 where  $\rho$  was only 0.58 makes clear that individual assessments of body fat cannot be derived from the BMI.

#### Measurement accuracy of B-mode US in thick SAT layers

In conventional US systems, the speed of sound used to calculate distances is 1540 m/s (“mean value for soft tissue”). Here, a sound speed of 1450 m/s was set for analysis of SAT in the evaluation software to avoid a sound speed error of about 6% (Herman 2007; Müller et al. 2016). The axial resolution in a US image depends on

pulse length, and the lateral resolution is determined by diffraction. Technically obtainable accuracy of thickness measurements equals approximately the wavelength (wavelength determines diffraction of US waves and, thus, the limits of image resolution). A probe frequency of 18 MHz results in an image resolution of about 0.1 mm, and 6 MHz results in approximately 0.3 mm. SAT borders are furrowed; therefore, biologically given limitations determine the obtainable accuracy. The influence of these biological limitations is minimized because the image evaluation algorithm used takes mean values of many thickness measurements in a given image.

In the overweight and obesity groups, low frequencies were necessary to penetrate the thick layers. In these groups, SAT thicknesses can be many centimeters, and in such cases the relative errors  $\delta_{\text{rel}}$  ( $= \delta/d$ , in %) are of relevance rather than the absolute errors  $\delta$  (in mm). For instance, at a 60-mm-thick SAT layer, the relative error (at 6 MHz) caused by border detection inaccuracy is approximately 0.3/60, that is, 0.5%. Such small border detection errors can be neglected compared with deviations because of wrong sound speed setting. A deviation of only 15 m/s (1%) would already affect a thickness measurement error of 1%. The absolute accuracy for *in vivo* measurements of thin fat layers using 12–18 MHz cannot be overcome by any other measurement technique. In thick layers, the correct choice of sound speed (1450 m/s) predominantly determines the accuracy. SAT is very sensitive to compression (Müller et al. 2013a), and errors caused by compression were minimized by using a thick layer of gel between the probe and the skin (Müller et al. 2013a, 2016).

#### Reliability of US SAT thickness measurements

Even breathing influences the SAT thickness. US images were therefore captured when the participant stopped breathing at midtidal expiration. Because of viscoelasticity, it is paramount to mark the sites and capture the images in the standardized positions (Müller et al. 2016). These authors had applied the standardized approach to analyze inter-observer reliability in a group of lean athletes (BMIs of 18.6–26.6 kgm<sup>-2</sup> and  $D_{\text{INCL}}$  of 10.2–51.2 mm); the median measurement deviation had been 0.24 mm, and the relative deviation had been 1.0%; this is close to the value of 1.1% found here in the intra-observer study for the lean group G1 (where  $D_{\text{INCL}}$  ranged from 12.5–77.4 mm).

In the study described here, the reliability analysis was extended to a group of participants with BMI values up to 40.3 kgm<sup>-2</sup> ( $D_{\text{INCL}}$  range, 12.5–244.9 mm). Intra-observer reliability of the measurements of two observers who investigated a subgroup of lean participants (G1, observer I) and a subgroup of overweight or obese people (G2, observer II) was analyzed. Absolute deviations (in

mm) between sums of thicknesses at the eight sites increased with thicker SAT layers (Figs. 3 and 4a, b); however, relative deviations were smaller in the overweight/obese group (median of  $ABS(\Delta_{INCL,rel}) = 0.5\%$ ) compared with the lean group (median  $ABS(\Delta_{INCL,rel}) = 1.1\%$ ) (Fig. 4c, d). This is not surprising when relative deviations are determined ( $\Delta_{rel} = 100 \cdot (\Delta/D_{MEAN})$ ) because tissue border detection errors play a larger role in lean persons in whom the sum of fat thicknesses  $D$  is small.

According to the standardized protocol (Müller *et al.* 2016), we used the new measurement site LT instead of the previously used EO. In some persons with thick fat layers in the abdominal region, it was difficult, and in some cases impossible, to mark the site EO because of skin and fat folds. LT provides interesting information on fat patterning. There was a significant difference ( $p < 0.01$ ) between men ( $N = 26$ ) and women ( $N = 12$ ). In men, the average contribution of the thickness at LT ( $d_{INCL,LT}$ ) to the sum of the eight sites ( $D_{INCL}$ ) was 15.1% (min 3.2%, max 25.7%); in women, it was 25.5% (min 17.9%, max 35.1%). This difference of about 70% between men and women indicates that LT represents an interesting fat depot area for comparative studies of SAT patterning in women and men.

A comparison of the measurement deviations ( $ABS(\delta)$ ) at the eight individual sites (Fig. 5a, b; Table 3) reveals that highest absolute deviations occurred at UA, LA and LT, where average SAT thicknesses were highest (Fig. 6a). This is to be expected: the plasticity, viscoelasticity and compressibility of subcutaneous adipose tissue cause the largest deviations from measurement to measurement at sites where SAT layers are thick. In the abdominal region, additional compression artifacts can occur because of breathing (the pressure status of the lungs when breathing is stopped at “midtidal” expiration can slightly differ from measurement to measurement). Fat thickness differences in the abdomen caused by the pressure change associated with the heartbeat can sometimes also be observed, but these effects are usually of minor importance.

The thickness of the SAT layer at LT may change when the leg is not positioned in the same way. At FT, measurement deviations were small (Fig. 5a). SAT compressibility is lowest at FT compared with the other US sites (Müller *et al.* 2013a). In the US image, quite often thick intermediate fasciae can be seen at FT, which may be the reason why skinfold measurement is difficult at this site.

The deviations when measurements were repeated (caused mainly by marking deviations, body position differences and intra-abdominal pressure changes) increased at all eight sites with increasing SAT thickness (Table 3). To obtain high reproducibility, it is very impor-

tant to position the person before the US measurement takes place precisely in accordance with the standardized US measurement technique (Müller *et al.* 2016).

Figure 6a illustrates the median values of SAT thicknesses at the individual sites. The ratio of the median deviation at a given site to the median thickness at this site is illustrated in Figure 6b. The relative errors (in %) were smallest at UA, LA and LT, where the average SAT thicknesses were largest (Table 3). Lowest absolute values (in mm) of measurement deviations occurred at erector spinae, distal triceps, brachioradialis, FT and medial calf where SAT depth was low.

In addition to investigating body composition problems in underweight persons and athletes (Byrne and McLean 2001; Nattiv *et al.* 2007; Sundgot-Borgen *et al.* 2013), results obtained here encourage to use this standardized US technique also for studies in groups of overweight and obese persons when highly accurate and reliable measurements of SAT are desired.

#### *Limitations of the study and future perspectives*

As this is a new measurement technique, it will take some time until extensive sets of reference data for low-weight, normal-weight, overweight and obese persons will be available. Validation studies against a four-component model (Ackland *et al.* 2012) will reveal which US sites will enable the best prediction of total body fat in particular groups of persons. Ethnicity, anthropometric parameters, sex and age are assumed to play a role in obtaining equations that fit optimally to the four-component model reference data. Such a validation study is in progress. The core aim of this study was to analyze reliability of the method; groups and numbers of participants were chosen for this purpose. A systematic comparative study of SAT patterning in various groups necessitates much larger and randomly chosen groups of participants. Comparative results described here are therefore only examples indicating the potential of the method for fat patterning studies.

## CONCLUSIONS

The standardized US measurement approach can be applied to all persons ranging from extremely lean to obese. Relative errors of US measurements groups with thick fat layers were even lower compared with groups with thin fat layers.

The eight sites represent trunk, arms and legs and accommodate inter-individual differences in SAT patterning. Also in obese persons, site marking is easy because distances are percentages of body height and only two anatomic landmarks are necessary (Müller *et al.* 2016). The site LT, where differences between

women and men were most distinct, was used here for the first time.

All eight sites overlie a muscle with a clearly visible fascia; this eases the acquisition of US images. In obese persons, a lower US frequency (*e.g.* 6 MHz) is necessary to detect the lower border (muscle fascia) of the subcutaneous fat.

Body composition disturbances, rapid weight changes, overweight, obesity and underweight are reasons for a wide range of severe medical problems. The US method enables cross-sectional and longitudinal studies, as well as individual fat patterning measurements on a scale much finer than available before in all persons ranging from extremely lean to obese. Although this US method is the laboratory reference for *in vivo* thickness measurements of uncompressed SAT (the US measurement accuracy has reached the biologically given limitations, and reliability has been maximized by standardizing the technique), it can also be applied as a field method because excellent US systems as small as notebooks are available today. Prices of US systems have rapidly decreased over the last years. It can be expected that this method will spread out into many fields of medicine in the near future.

Guided training is recommended to reach the accuracy and reliability levels that are attainable with this US method when applied in standardized manner (Müller *et al.* 2016).

## REFERENCES

- Ackland TR, Lohman TG, Sundgot-Borgen J, Maughan RJ, Meyer NL, Stewart AD, Müller W. Current status of body composition assessment in sport: Review and position statement on behalf of the Ad Hoc Research Working Group on Body Composition Health and Performance, under the auspices of the IOC. Medical Commission. *Sports Med* 2012;42:227–249.
- Both RA, Goddard BA, Paton A. Measurement of fat thickness in man: A comparison of ultrasound, Harpenden calipers and electrical conductivity. *Br J Nutr* 1966;20:719–725.
- Bullen BA, Quaade F, Olessen E, Lund SA. Ultrasonic reflections used for measuring subcutaneous fat in humans. *Hum Biol* 1965;37:375–384.
- Byrne S, McLean N. Eating disorders in athletes: A review of the literature. *J Sci Med Sport* 2001;4:145–159.
- Herman IP. *Physics of the human body*. Berlin: Springer-Verlag; 2007.
- Heymsfield SB, Lohman TG, Wang ZM, Going SB. *Human body composition*. 2nd edition. Champaign, IL: Human Kinetics 2005.
- Hill JO, Wyatt HR, Peters JC. The importance of energy balance. *US Endocrinol* 2013;9:27–31.
- Müller W. Determinants of ski-jump performance and implications for health, safety and fairness. *Sports Med* 2009a;39:85–106.
- Müller W. Towards research-based approaches for solving body composition problems in sports: Ski jumping as a heuristic example. *Br J Sports Med* 2009b;43:1013–1019.
- Müller W, Gröschl W, Müller R, Sudi K. Underweight in ski jumping: The solution of the problem. *Int J Sports Med* 2006;27:926–934.
- Müller W, Horn M, Fürhapter-Rieger A, Kainz P, Kröpfl JM, Ackland TR, Lohman TG, Maughan RJ, Meyer NL, Sundgot-Borgen J, Stewart AD, Ahammer H. Body composition in sport: Interobserver reliability of a novel ultrasound measure of subcutaneous fat tissue. *Br J Sports Med* 2013b;47:1036–1043.
- Müller W, Horn M, Fürhapter-Rieger A, Kainz P, Kröpfl JM, Maughan RJ. 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* 2013a;47:1028–1035.
- Müller W, Lohman TG, Stewart AD, Maughan RJ, Meyer NL, Sardinha LB, Kirihennedige N, Reguant-Closa A, Risoul-Salas V, Sundgot-Borgen J, Ahammer H, Anderhuber F, Fürhapter-Rieger A, Kainz P, Materna W, Pilsil U, Pirstinger W, Ackland TR. 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.
- Müller W, Maughan RJ. The need for a novel approach to measure body composition: Is ultrasound an answer? *Br J Sports Med* 2013;47:1001–1002.
- Nattiv A, Loucks AB, Manore MM, Sanborn CF, Sundgot-Borgen J, Warren MP. American College of Sports Medicine position stand: The female athlete triad. *Med Sci Sports Exerc* 2007;39:1867–1882.
- Stewart AD, Marfell-Jones M, Olds T, Deridder H. *International standards for anthropometric assessment*. Lower Hutt, New Zealand: International Society for the Advancement of Kinanthropometry; 2011.
- Sundgot-Borgen J, Meyer NL, Lohman TG, Ackland TR, Maughan RJ, Stewart AD, Müller W. How to minimise the health risks to athletes who compete in weight-sensitive sports review and position statement on behalf of the Ad Hoc Research Working Group on Body Composition, Health and Performance, under the auspices of the IOC Medical Commission. *Br J Sports Med* 2013;47:1012–1022.
- World Health Organization (WHO). *Global status report on noncommunicable diseases 2014*. Geneva: Author; 2015.
- World Health Organization (WHO) Expert Committee. *Physical status: The use and interpretation of anthropometry*. Geneva: World Health Organization; 1995.

## Appendix

### ABBREVIATIONS:

SAT	Subcutaneous adipose tissue
US	Ultrasound
ROI	Region of interest
G1, G2	Group1 (lean), Group2 (overweight/obese)
M1, M2, M3	Measurement series

### Parameters and variables:

$m$	Body mass, in kg
$h$	Stature, in m
$s$	Sitting height, in m
BMI	Body mass index: $BMI=m/h^2$ , in $kgm^{-2}$
$MI_1$	Mass index: $MI_1=0.53 \cdot m/(hs)$ , in $kgm^{-2}$
$d$	SAT thickness at a given site, in mm (this is the average of the distances measured within the region of interest)
$D$	Sum of SAT-thicknesses at all eight sites in a given participant, in mm
$\delta$	measurement deviation of each measurement (M1, M2, M3) from the mean of the three measurements at a given site in a given subject, in mm
$\delta_{rel}$	$\delta_{rel}= 100 \cdot \delta/d_{MEDIAN}$ , in [%]
$\Delta$	Deviation of the sum of eight sites from the mean of the sums of the three measurements in a given participant, in mm
$\Delta_{rel}$	$\Delta_{rel}= 100 \cdot \Delta/D_{MEAN}$ , in [%]
EXCL	Excluded; indicates that the fibrous structures embedded in the SAT are not included in the thickness value.
INCL	Included; indicates that the fibrous structures are included

**Statistics:**

N	Number of values
ABS	Absolute value of a number
MEAN	Mean value
SD	Standard deviation
SEE	Standard error of estimate
$\rho$	Spearman's rank correlation coefficient (Spearman's rho)

**US Sites:**

UA	Upper abdomen
LA	Lower abdomen
ES	Erector spinae
DT	Distal triceps
BR	Brachioradialis
LT	Lateral thigh
FT	Front thigh
MC	Medial calf
EO	External oblique (optional)