Pre Print Article:

Relationships and comparative reliability of ultrasound derived measures of upper and lower limb muscle thickness, and estimates of muscle area from anthropometric measures

Emily Budzynski-Seymour [1], James Fisher [1], Jürgen Giessing[2], Paulo Gentil[3], James Steele [1,4]

[1] School of Sport, Health, and Social Sciences, Solent University, Southampton, United Kingdom
[2] University of Koblenz-Landau,, Landau, Germany
[3] Federal University of Goias, Goiania, Brazil
[4] ukactive Research Institute, London, United Kingdom

Contact for correspondence:

Email: emily.budzynski-seymour@solent.ac.uk james.steele@solent.ac.uk

Twitter: @Emily_CBS @JamesSteeleII

DOI: 10.31236/osf.io/ujktq

Citation: Budzynski-Seymour, E, Fisher, J, Giessing,J, Genil,P, and Steele,J. 2019. Relationship and comparative reliability of ultrasound derived measures of upper and lower limb muscle thickness, and estimates of muscle area from anthropometric measures.

All authors agree to share this work

Abstract: The gold standard measure for assessing muscular size currently is magnetic resonance imaging; however, it is expensive and not easily accessible. Both anthropometric techniques (AN) and ultrasound (UT) are commonly employed methods to measure muscle size. However, the degree to which these approaches offer similar information has not been examined. The aim of the study was to investigate the relationship between UT and AN measurements of muscle thickness in addition to their comparative reliability. Fifteen males (27±9 years) volunteered to take part in the study and underwent both AN and UT measures, taken to assess their upper arm and upper leg muscle size on separate days a week apart. Correlations between the two measures ranged from r=0.548-0.918 (p<0.05) suggesting a good relationship and thus comparable information. Results showed similar coefficient of variation (CV%) for the upper leg (AN 2.3%, UT 2.4%), but slightly greater reliability for UT results for the upper arm (AN 5.5%, UT 2.8%). It appears that both methods are reliable approaches to measurement of muscle size, though AN likely represents a lower cost and greater ease of use. Researchers should consider this when deciding upon which approach to use in the assessment of muscle size in the absence of gold standard approaches.

Keywords: Muscle size, Muscle thickness, Ultrasound, Anthropometry, Reliability

1. Introduction

To consider the degree of hypertrophy (i.e. enlargement of muscular size), the cross-sectional area of the muscle is often measured in both a clinical and non-clinical settings. Clinically, measurements of muscular size have been adopted to evaluate the nutritional status of a variety of populations with muscle wasting diseases or injuries, and in non-clinical settings it can be used to assess the effectiveness of exercise interventions (1). The assessment of muscular size can be achieved through a variety of different methods including: bioelectrical impedance (BIA), dual energy X-ray (DXA), computed tomography (CT), magnetic resonance imaging (MRI), and B mode ultrasound (2).

BIA, which is a popular method for estimating body composition, uses multifrequency electrical charges to estimate the extracellular and intracellular fluids (3). BIA often provides a whole body estimation (though newer devices can provide regional estimates), however DXA can be used to measure both whole body and regional composition. A DXA scan can determine the relative levels of bone free mineral free lean tissue, fat, and bone mineral in all the pixels of the area being scanned (4). Dual energy X-ray does exposure the subject to levels relatively minor levels of radiation, though in contrast CT produces higher levels of radiation exposure. CT is a method which provides three-dimensional views of the bodies organs, and bone density measurements, however it has lost popularity due to the high levels of radiation exposure (5).

MRI is a sophisticated laboratory technique that has been shown to be both valid and reliable, and is particularly useful at analysing skeletal muscles as it differentiates clearly between muscle, fat, ligaments and vessels all visible on the scan (1). It is considered the gold standard of skeletal muscle imaging (6). Although MRI is an effective method it requires expensive equipment, this means it is not easily accessible or feasible for most practitioners.

Anthropometry is a simpler method for measurement of bodily proportions that is inexpensive and non-invasive (7). The instruments commonly used for anthropometry are portable and require minimal training (8). Research using anthropometry to measure muscle cross sectional area has shown a good agreement with MRI (r^2=0.98, p<0.001) in clinical populations (9). Indeed, equations have been derived from studies using MRI for use with anthropometric measures (e.g. circumference and skinfold thickness measures) for estimation of muscle size for a range of populations (8,1, 9). For these reasons it has become more popular in research and is now commonly used for estimating muscular size as seen in a number of recent studies (10, 11, 12).

Another method which recently many studies have also begun to more commonly employ is ultrasound to measure muscle size by measuring muscle thickness (13). Similarly, to anthropometric methods, ultrasound can be easily applied in both clinical and field-based research and is non-invasive, cheap, quick and a safe imaging technique (14). It has been shown to be a reliable tool for assessing skeletal muscle thickness in various populations (ICC=0.99) (6). However, though considerably cheaper than MRI potentially explaining its increased use in research, ultrasound still presents a greater cost barrier to use than anthropometry (15).

Despite MRI being the gold standard method for measurements of muscle thickness it can present difficulties with regards to cost and accessibility. Both anthropometry and ultrasound approaches have become more common in research examining muscle size, particularly in response to resistance training. However, the relationship between these two measures is not known, nor their comparative reliability. Such information might help in appraising the literature utilising these techniques and inform future research designs. Therefore, the aim of this research is to investigate the relationship between ultrasound and anthropometric measurements of muscle size and their comparative reliability.

2. Materials and Methods

2.1. Participants

Fifteen recreationally active healthy adult males (age = 27 ± 9 years) volunteered to take part in this study. They were recruited using convenience sampling. They all provided written informed consent prior to any testing, and the study was approved by the Health, Exercise and Sport Science ethics committee at the lead author's institution.

2.2. Procedures

The participants were required to attend two testing days, the same time each day with exactly one week between the two sessions. They were instructed to avoid strenuous physical activity or exercise 3-5 days prior to the assessments, and to maintain normal hydration. A crossover study design was used as all participants underwent both ultrasound (UT) and anthropometric (AN) measurements. These were taken at the thigh and upper arm on their right side to determine muscle thickness and estimate muscle area respectively. One single measure was taken at each site, all of the measures were taken by the same researcher with the participants in the same position and the same sites for both the UT and AN measures. The researcher conducting the measures was experienced in use of both approaches.

Harpenden Callipers (Harpenden, Baty International, UK) were used for skinfolds, and circumference measures were taken using an anthropometry tape measure (Seca Tape Measure, Seca, UK). The mid-thigh circumference was taken midway between the inguinal crease and the proximal border of the patella with the participant standing on an anthropometry box. The mid arm circumference and skinfold were taken midway between the tip of the acromion and olecranon process, with the participant standing with their elbow extended and forearm supinated. To calculate the thigh muscle area estimate the following equation was used: (4.68 X mid-thigh circumference) - (2.09 X anterior thigh skinfold) – 80.99 (1). For total arm muscle area estimate the following equation was used: (Mid arm circumference - π X Tricep skinfold thickness)^2/4 π, (7).

For the UT measures the same sites were used and with the participant in the same position in order to maintain parity between comparison of the two approaches. Muscle thicknesses were measured using an M7 Diagnostic Ultrasound System (Shenzhen Mindray Bio-Medical Electronics Co. Ltd., China) and B-mode ultrasound imaging. A 7.5MHz linear array transducer was used with a scanning head coated with a soluble transmission gel to improve the imaging quality. A single transverse image was taken at each site once and then digital callipers in the software used to measure muscle thickness of the elbow flexors, elbow extensors, and quadriceps as the distance between the subcutaneous adipose tissue-muscle interface and the muscle-bone interface. Figure 1 shows an example image.

Figure 1: Example UT image of participant quadricep thickness measurement.

2.3. Statistical analysis

For the AN measures the variables were arm muscle area (AMA), and thigh muscle area (TMA), and the UT variables were elbow flexor thickness (FT), elbow extensor thickness (ET), arm total thickness (FT+ET; AT) and quadriceps thickness (QT). Using the Statistical Package for Social Science (Version 22.0) with significance accepted at $p<.05$, a Shapiro-Wilk test was used to test for normality of distribution. As all data was shown to be normally distributed, Pearson's correlations were performed to examine the associations between the first and second days of testing. Correlation coefficients were interpreted as weak (r = 0.30 – 0.50), moderate (r = 0.51 – 0.70), or strong (r > 0.70). Between day reliability across the two testing days was also calculated for each method and comparison of the relative errors between methods made using the coefficient of variance (CV%) using existing spreadsheets for their calculation ().

3. Results

For day 1, strong significant correlations were found for AMA and ET (r=0.859 p<0.01), AMA and AT (r=0.918 p<0.01), and TMA and QT (r=0.918 p<0.01), with a moderate significant correlation found between AMA and FT (r=0.642 p<0.05). For day 2, strong significant correlations were found for AMA and AT (r=0.871 p<0.01) and TMA and QT (r=0.831 p<0.01), and moderate significant correlations were found for both AMA and FT (r=0.548 p<0.05) and AMA and ET (r=0.622 p<0.01).

Table 1: Correlations between AN and UT measures		
Day 1:	Arm Muscle Area (AN)	Thigh Muscle Area (AN)
Flexor (UT)	0.642**	-
Extensor (UT)	0.859**	-
Arm total (UT)	0.918**	-
Quad (UT)	-	0.918**
Day 2:	Arm Muscle Area (AN)	Thigh Muscle Area (AN)
Flexor (UT)	0.548*	-
Extensor (UT)	0.622**	-
Arm total (UT)	0.871**	-
Quad (UT)	-	0.831

*Significance p<0.05

**Significance p<0.01

A larger coefficient of variance was reported for the AMA using the AN measures when compared to UT (5.5% AN, 2.8% UT) suggesting a better reliability score when using the ultrasound. There was only a small difference between the coefficient of variance for the quadriceps and thigh (2.3% AN, 2.4% UT) suggesting similar reliability scores.

Table 2: Means, SD and CV (%) for test-retest across days 1 and 2.

	Upper Arm				Thigh	
	Arm Muscle Area (cm²)	Arm Total Thickness (cm)	Extensor Thickness (cm)	Flexor Thickness (cm)	Thigh Muscle Area (cm²)	Quadriceps Thickness (cm)
Mean	71.9	8.0	4.4	3.6	147.4	5.6
SD	12.0	0.8	0.6	0.5	17.2	0.8
CV (%) [95%CIs]	5.5 [3.9, 9.2]	2.8 [2.0, 4.7	5.3 [3.7, 8.8]	4.9 [3.5, 8.2]	2.3 [1.6, 4.0]	2.4 [1.7, 4.4]

4. Discussion

The aim of this research was to investigate the relationship between ultrasound and anthropometric measurements of muscle size in addition to their comparative reliability. Correlations were moderate to strong between AN and UT measures for both days 1 and 2. Furthermore, both methods have minimal variation with the UT showing slightly better reliability compared with AN methods for the upper arm. These results suggest both are reliable tools to measure muscle thickness, and that they both provide similar, or at least related, measures of muscular size.

The correlations reported between methods and days showed either strong or moderate relationships between AN and UT measures of muscle size. It appeared that the relationships for the upper arm were strongest when considering the AT thickness using UT. As the AN measures used estimate total upper AMA this is unsurprising. However, though it suggests that AN estimates of upper arm muscle size may be related to those obtained using UT, caution should be employed in drawing inferences from AN based estimates of AMA to infer muscle size for the elbow extensors or flexors alone. Contrastingly, there were strong relationships between UT measures of QT and AN estimates of TMA. Upon retrospective analysis of the data using equations from Housh *et al.*, (1995) for quadriceps CSA alone we found correlations of r=0.808-0.904 ($p \leq 0.001$) with UT measured QT. As such, it may be that AN estimates of TMA are more representative of quadriceps muscle size perhaps due to the relatively greater proportion of the quadriceps TMA that they contribute to. These relationships are in line with previous research considering body composition which has reported that UT and AN methods are moderately correlated ($r > 0.68$, $p < 0.01$) (16).

Although the CVs reported suggest that, compared with AN derived estimates of AMA, UT measures of AT thickness were slightly more reliable for the upper arm, the AN derived estimate of TMA and UT measures of QT results showed similar CVs suggesting that both are comparatively reliable methods of assessment. Both UT and AN measures have been seperately reported to be reliable in a range of populations both sporting and clinical (7,1,17,16). Indeed, our results suggest that the two offer largely comparable reliability with respect to estimation of appendicular muscle size. This is similar to that reported by Wagner *et al.*, (16) for body composition whereby, though UT measures were slightly more reliable, both UT and AN offered excellent reliability.

These results show the strong relationships and comparative reliability of the two measurement types. As such, one does not appear to be inherently superior to the other. As a result, the data suggests that both might be recommended, though considering the relatively lower cost AN might be considered more accessible for both researchers and practitioners. Although both measures are relatively inexpensive and simple to administrate AN measures are more so; and therefore more accessible for practitioners in particular (15). This may be where the implementation of AN may be superior over the use UT. Research has found that AN measures are comparable to gold standard measures using MRI (r^2=0.98, p<0.001), supporting its continued use (9). Wade and Gorgey (9) aimed to identify an accurate and affordable method to quantify muscle size following a spinal cord injury, they compared AN measures and the gold standard MRI. They concluded that AN measurement of muscle cross sectional area showed good agreement with MRI and therefore its use is justified in research, specifically in a clinical setting. The equations used here with AN measures for estimation of both upper arm and thigh muscle size have also been validated in healthy and undernourished chronically ill participants (7, 1). Furthermore, in a performance environment AN measures are commonly used in numerous sports, this could be associated with its ease of use and reliable results. Examples of its use in a performance setting include: to track the effects of plyometric training (18), monitor the effects if resistance training (19), as a follow up measure looking at aerobic power in older track athletes (20) and in investigations into strength and power development (21). Indeed, DeFreitas et al. (22) have shown that changes in muscle size can be monitored similarly using AN methods even compared to CT.

However, the present data also revealed that the UT was slightly more reliable when measuring the upper arm when compared to AN methods and thus, if accessible, this approach may be more appealing to use. Previous research has found strong correlations (r=0.891-0.946) between UT muscle thickness measures in the forearm arm compared to the gold standard measure of MRI (2). AN measurement of muscle size involves estimation through the use of equations, and the use of dated equations in research has revealed a 20-25% overestimation of AMA (7). However, more recent and altered equations that are used in current research reduce this over estimation and increase their overall accuracy (23). This suggests that AN measures may still be a reliable tool for measuring AMA if the most recent equations are used in the calculations.

Our results suggest that AN and UT measures of muscle size are strongly related and comparably reliable and thus may offer value in the assessment of muscle size and potentially hypertrophic changes as a result of an intervention. However, a limitation of the current research was that measures were only taken of the associations between two individual time points and not in the changes seen from one to another. Recent research has suggested that, not only do UT and MRI measures of muscle size correlated strongly at a single time point, changes in muscle size after a resistance training intervention measured using both of these methods also correlate well (6). Considering that both UT and AN based measured of muscle size correlated with MRI at single time points, and that their reliability is comparable, it might be assumed that changes in these outcomes might also correlate with one another. Indeed, as noted AN methods have at least been shown to measure changes comparably to CT (22). However, at present whether AN methods measure similar changes in comparison to UT, or gold standard measures using MRI, has not been investigated. Therefore, it is suggested that future research should look at the relationship between changes in both UT and AN measures of muscle size as a result of a resistance training intervention.

5. Conclusions

This study has shown the benefits of using both AN and UT measures to assess muscle size and demonstrated their strong relationships and comparable reliability. Both measures provide similar CV scores and there are moderate to strong correlations between the two measurement types. As such, both might offer similar information regarding muscle size. However, reliability may be slightly greater for UT based measures, though AN based measures likely represent lower costs and greater ease of use. Researchers and practitioners should consider this when deciding upon which approach to use in the assessment of muscle size in the absence of gold standard approaches.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Housh, D. J., Housh, T. J., Weir, J. P., Weir, L. L., Johnson, G. O., & Stout, J. R. Anthropometric estimation of thigh muscle cross-sectional area. *Medicine & Science in Sports & Exercise* (1995); **27(**5): 784-791

2. Abe T, Loenneke JP, Thiebaud RS. The use of ultrasound for the estimation of muscle mass: one site fits most? *J Cachexia Sarcopenia Muscle* (2018); **9**(1): 213-214

3. Völgyi, E., Tylavsky, F. A., Lyytikäinen, A., Suominen, H., Alén, M., & Cheng, S. (2008). Assessing body composition with DXA and bioimpedance: effects of obesity, physical activity, and age. *Obesity*, *16*(3), 700-705.

4. LaForgia, J., Dollman, J., Dale, M. J., Withers, R. T., & Hill, A. M. (2009). Validation of DXA body composition estimates in obese men and women. *Obesity*, *17*(4), 821-826.

5. Kabayel, D. D. (2016). The role of quantitative computed tomography and magnetic resonance imaging in diagnosis and follow-up of osteoporosis: A review. *Turkish Journal of Physical Medicine & Rehabilitation/Turkiye Fiziksel Tip ve Rehabilitasyon Dergisi*, *62*(3).

6. Franchi, M. V., Longo, S., Mallinson, J., Quinlan, J. I., Taylor, T., Greenhaff, P. L., & Narici, M. V. Muscle thickness correlates to muscle cross-sectional area in the assessment of strength training-induced hypertrophy. *Scandinavian journal of medicine & science in sports* (2018); **28**(3): 846-853

7. Heymsfield, S. B., McManus, C., Smith, J., Stevens, V., & Nixon, D. W. Anthropometric measurement of muscle mass: revised equations for calculating bone-free arm muscle area. *The American Journal of Clinical Nutrition* (1982); **36**(4): 680-690

8. Lee, R. C., Wang, Z., Heo, M., Ross, R., Janssen, I., & Heymsfield, S. B. Total-body skeletal muscle mass: development and cross-validation of anthropometric prediction models–. *The American journal of clinical nutrition* (2000); **72**(3): 796-803

9. Wade, R. C., & Gorgey, A. S. Anthropometric prediction of skeletal muscle cross-sectional area in persons with spinal cord injury. *Journal of Applied Physiology* (2017); **122**(5): 1255-1261.

10. de França HS, Branco PA, Guedes Junior DP, Gentil P, Steele J, Teixeira CV. The effects of adding single-joint exercises to a multi-joint exercise resistance training program on upper body muscle strength and size in trained men. *Appl Physiol Nutr Metab* (2015); **40**(8): 822-826

11. Gentil, P., Soares, S., & Bottaro, M. Single vs. Multi-Joint Resistance Exercises: Effects on Muscle Strength and Hypertrophy. *Asian J Sports Med* (2015); **6**(2): e24057

12. Gentil, P., Soares, S. R., Pereira, M. C., da Cunha, R. R., Martorelli, S. S., Martorelli, A. S., & Bottaro, M. Effect of adding single-joint exercises to a multi-joint exercise resistance-training program on strength and hypertrophy in untrained subjects. *Appl Physiol Nutr Metab* (2013); **38**(3): 341-344

13. Bemben MG. Use of diagnostic ultrasound for assessing muscle size. *J Strength Cond Res* (2002); **16**(1): 103-108

14. Abe T, Loenneke JP, Thiebaud RS. Morphological and functional relationships with ultrasound measuredmuscle thickness of the lower extremity: a brief review. *Ultrasound* (2015); **23**(3): 166-173

15. Selkow, N. M., Pietrosimone, B. G., & Saliba, S. A. Subcutaneous thigh fat assessment: a comparison of skinfold calipers and ultrasound imaging. *Journal of Athletic Training* (2011); **46**(1): 50-54.

16. Wagner, D. R., Cain, D. L., & Clark, N. W. Validity and reliability of A-mode ultrasound for body composition assessment of NCAA division I athletes. *PloS one (2016)*; **11(**4): e0153146.

17. Tillquist, M., Kutsogiannis, D. J., Wischmeyer, P. E., Kummerlen, C., Leung, R., Stollery, D., ... & Heyland, D. K. Bedside ultrasound is a practical and reliable measurement tool for assessing quadriceps muscle layer thickness. *Journal of Parenteral and Enteral Nutrition* (2014); **38**(7): 886-890.

18. Chelly, M. S., Ghenem, M. A., Abid, K., Hermassi, S., Tabka, Z., & Shephard, R. J. Effects of in-season short-term plyometric training program on leg power, jump-and sprint performance of soccer players. *The Journal of Strength & Conditioning Research* (2010); **24**(10): 2670-2676

19. Christou, M., Smilios, I., Sotiropoulos, K., & Volaklis, K. Effects of resistance training on the physical capacities of adolescent soccer players. *Journal of Strength & Conditioning Research* (2006); **20**(4): 783

20. Pollock, M. L., Mengelkoch, L. J., Graves, J. E., Lowenthal, D. T., Limacher, M. C., Foster, C., & Wilmore, J. H. Twenty-year follow-up of aerobic power and body composition of older track athletes. *Journal of Applied Physiology* (1997); **82**(5): 1508-1516

21. Chtara, M., Chaouachi, A., Levin, G. T., Chaouachi, M., Chamari, K., Amri, M., & Laursen, P. B. Effect of concurrent endurance and circuit resistance training sequence on muscular strength and power development. *The Journal of Strength & Conditioning Research* (2008); **22**(4): 1037-1045

22. DeFreitas, J. M., Beck, T. W., Stock, M. S., Dillon, M. A., Sherk, V. D., Stout, J. R., & Cramer, J. T. A comparison of techniques for estimating training-induced changes in muscle cross-sectional area. *Journal of Strength and Conditioning Research* (2010); **24**(9): 2383-2389

23. Heymsfield, S. B., Gonzalez, M. C., Lu, J., Jia, G., & Zheng, J. Skeletal muscle mass and quality: evolution of modern measurement concepts in the context of sarcopenia. *Proceedings of the Nutrition Society* (2015); **74**(4): 355-366