Bedside Ultrasound Is a Practical and Reliable Measurement Tool for Assessing Quadriceps Muscle Layer Thickness

Maggie Tillquist, MD[1], **Demetrios J. Kutsogiannis, MD, MHS, FRCPC**[2], **Paul E. Wischmeyer, MD**[3], **Christine Kummerlen, MD**[4], **Roger Leung, BSc**[5], **Daniel Stollery, MD, FCCP, FRCPC**[2], **Constantine J. Karvellas, MD, FRCPC**[2], **Jean-Charles Preiser, MD, PhD**[6], **Nora Bird, VMD, BS, RDMS**[7], **Rosemary Kozar, MD, PhD**[7], and **Daren K. Heyland, MD, FRCPC, MSc**[5]

[1]Brigham and Women's Hospital, Boston, MA, USA

[2]University of Alberta, Edmonton, AB, Canada

[3]University of Colorado School of Medicine, Aurora, CO, USA

[4]University Hospital, University of Strasbourg, Strasbourg, France

[5]Kingston General Hospital, Kingston, ON, Canada

[6]Erasme University Hospital, Brussels, Belgium

[7]UT Health Science Center of Houston, Houston, TX, USA

Abstract

Background—Critically ill patients commonly experience skeletal muscle wasting that may predict clinical outcome. Ultrasound is a noninvasive method that can measure muscle quadriceps muscle layer thickness (QMLT) and subsequently lean body mass (LBM) at the bedside. However, currently the reliability of these measurements are unknown. The objectives of this study were to evaluate the intra- and interreliability of measuring QMLT using bedside ultrasound.

Methods—Ultrasound measurements of QMLT were conducted at 7 centers on healthy volunteers. Trainers were instructed to perform measurements twice on each patient, and then a second trainee repeated the measurement. Intrarater reliability measured how consistently the same person measured the subject according to intraclass correlation (ICC). Interrater reliability measured how consistently trainer and trainee agreed when measuring the same subject according to the ICC.

Results—We collected 42 pairs of within operator measurements with an ICC of .98 and 78 pairs of trainer-to-trainee measurements with an ICC of .95. There were no statistically significant differences between the trainer and trainee results (trainer and trainee mean = −0.028 cm, 95% CI = −0.067 to −0.011, P = .1607).

Conclusions—Excellent intra- and interrater reliability for ultrasound measurements of QMLT in healthy volunteers was observed when performed by a range of providers with no prior

Corresponding Author: Daren Keith Heyland, MD, FRCPC, MSc, Kingston General Hospital, Angada 4 Room 5-416, 76 Stuart Street, Kingston, ON K7L 2V7, Canada. dkh2@queensu.ca.

ultrasound experience, including dietitians, nurses, physicians, and research assistants. This technique shows promise as a method to evaluate LBM status in ICU or hospital settings and as a method to assess the effects of nutrition and exercise-based interventions on muscle wasting.

Keywords

ultrasound; critical care; muscle layer thickness; muscle wasting

Clinical Relevancy Statement

Skeletal muscle weakness/wasting and consequent impaired physical function are common findings amongst patients in the intensive care unit (ICU). In critically ill patients, immobilization, sepsis, organ failure, and systemic inflammation are all associated with muscle wasting. This wasting is associated with ICU-acquired weakness, which is a consequence of critical illness, which can lead to impaired physical functions for years post-ICU discharge. Interventions, such as improved nutrition delivery, have been proposed to prevent or attenuate loss of lean body mass (LBM) and subsequent ICU-acquired weakness. However, to date clinically relevant bedside methods for assessing LBM have not been available. This article describes the use of a new bedside ultrasound technique to assess LBM, which shows promise for use in the ICU and hospital setting. Our data specifically demonstrate excellent intra- and interrater reliability can be obtained for ultrasound measurements in healthy volunteers when performed by a range of providers with no prior ultrasound experience, including dietitians, nurses, physicians, and research assistants. This technique shows promise as a method to evaluate the LBM status in the ICU or hospital setting by a range of providers and as a method to assess the effects of nutrition- and exercise-based interventions on muscle wasting.

Introduction

Skeletal muscle weakness and consequent impaired physical function are common findings amongst patients in the intensive care unit (ICU).[1,2] In critically ill patients, immobilization, sepsis, organ failure, and systemic inflammation are all associated with muscle wasting.[3–5] Investigators estimate that critical illness myopathy affects between 25% and 100% of ICU patients, depending on the assessment tool utilized and duration of stay in the ICU.[3] Furthermore, critical illness myopathy is an independent predictor of patient morbidity[1] and mortality[6] and long-term loss of functional autonomy.[1]

As ICU-acquired muscle weakness remains a clinical diagnosis, it is critical to investigate new methods to measure lean body mass (LBM) in the critically ill. A functional approach utilizes electromyogram and nerve conduction study, but the use of these techniques are limited by the low yield of data obtained from patients unable to participate.[3] In the ICU, the most commonly used clinical tool to assess muscle weakness is manual muscle testing (MMT) using the Medical Research Council muscle strength sum score.[4,7] Yet similar to other measurement tools, MMT is limited by the low yield of data obtained from patients that may have cognitive impairments or are otherwise unable to participate.[3]

Given the impact of ICU-acquired muscle weakness on clinical outcomes, recent research has focused on noninvasive methods measuring muscle thickness or cross-sectional area.[3] Initial efforts investigating anthropometric measurements are limited by their dependence on fluid state, and ultimately have not correlated with LBM measured by CT scanning, DEXA scanning in the critically ill. Recently, new research found that ultrasonographic measurements of the quadriceps muscle appears to be as accurate as CT scan based muscle mass analysis,[8] which is a standard along with DEXA scanning in the assessment of muscle mass.[9] Previous research shows assessment of two ultrasound measurements of the rectus femoris muscle thickness to be representative of overall muscle mass vs MRI muscle assessement.[10] Our study design to measure quadriceps muscle layer thickness (QMLT) is similar to that as reported by Campbell et al and Gruther et al.[11–12] Important to this techniques potential use in the ICU, the study by Gruther et al showed that ultrasound is a valid and practical measurement tool for documenting muscle mass (eg, muscle layer thickness) as part of the daily routine at an ICU. Furthermore, they showed that loss of muscle mass demonstrated a negative correlation with length of stay, and seemed to be higher during the first 2–3 weeks of immobilization/ICU stay.[12] In addition, ultrasound is easily applicable at the bedside, is available in most ICUs, and costs little compared with other techniques. Initial data suggest that ultrasound QMLT measurements may be the ideal tool for assessment of ICU-acquired skeletal muscle weakness.[3]

Before we can validate bedside U.S. assessments of LBM, we must demonstrate the intrarater and interrater reliability of ultrasonographic measurements. Thus, our aim was to assess intrarater and interrater reliability of bedside ultrasound to evaluate QMLT in multiple institutions across multiple assessors.

Study Design

We conducted a prospective observational study of measuring QMLT in healthy volunteers. The study involved a convenience sample of seven centers involved in critical care nutrition research with each center having a trainer and one or more trainees conducting the research. Although volunteers per study site varied, trainers were asked to measure a minimum of five healthy subjects recruited at each center. Centers were asked to recruit volunteer subjects from ICU staff members across a range of body mass indices whenever possible. There were no formal inclusion or exclusion criteria.

The QMLT protocol was created and approved by a registered diagnostic medical sonographer. QMLT was calculated by measuring at the border between the lower third and upper two-thirds between the anterior superior iliac spine (ASIS) and the upper pole of the patella, as well as the measurement of the midpoint between the ASIS and the upper pole of the patella. The right and left quadriceps value assessed was the average of these four readings over the right and left legs (two at each site) rather than the individual measurements (see Figure 1). To standardize measurements, a procedural manual and an accompanying training video were developed and sent to all sites. Trainers were physicians with advanced training in bedside ultrasound. Trainees were comprised of dietitians, nurses, physiotherapists, investigators, coordinators, and technicians, the majority with no prior ultrasound experience.

To standardize measurements, a procedural manual and an accompanying training video were developed and sent to all sites. QMLT was calculated by measuring at the border between the lower third and upper two-thirds between the ASIS and the upper pole of the patella, as well as the measurement of the midpoint between the ASIS and the upper pole of the patella. The right and left quadriceps value assessed was the average of these four readings over the right and left legs (two at each site) rather than the individual measurements (see Figure 1).

Our study evaluated both intrarater and interrater reliability. Intrarater reliability evaluates how consistently an individual trainer measure the same subject according to the intraclass correlation coefficient (ICC) ICC = *between-subject variance*/(*between-subject variance* + *within-subject variance*). Between-subject variance is the variance between different subjects measured by the same trainer. Within-subject variance refers to the variance between two measurements from an individual trainer on the same subject.

Interrater reliability measures the variance between trainer and trainee in their evaluation of the same subject. The reliability is calculated by the ICC as defined above, except that *within-subject variance* measures the variance between trainer and trainee in their evaluation of the same subject.

Trainers were asked to repeat measurements on the same leg and then have one or more trainees repeat the same measurements. The mean difference between the first and second trainer measurements as well as the first trainer and the trainee measurements and the measurements on the left and right sides were evaluated by the paired *t* test. All values were reported at the 95% confidence interval with significant differences being reflected by a *P* value of .05.

Results

Subject Demographics

The study included 78 healthy volunteers across seven centers, including multiple sites in Canada, the United States, Belgium, and France. Volunteer subject demographics are shown in Table 1.

In measurement of the QMLT, the first trainer measurement had a mean of 2.01 cm as an average between the left and right leg with a standard deviation of 0.52 cm. The second trainer measurement had a mean of 2.07 cm as an average between the left and right leg with a standard deviation of 0.50 cm. Finally, the trainee measurement had a mean of 2.09 cm as an average between the left and right leg with a standard deviation of 0.52 cm.

Analysis of Intra- and Interrater Reliability

As shown in Table 2, 42 pairs of within operator measurements (range across sites 0–12) were evaluated with an ICC of 0.98 (range across sites 0.94–>0.99). There was a small but statistically significant difference between the first and second measures by the trainer (mean = 0.033 cm, 95% CI = 0.004–0.061, *P* = .0247; Figure 2). There were 78 pairs of trainer to trainee measurements (range across sites 5–18) with an ICC of 0.95 (range across

sites 0.73–>0.99; Table 3). There was only one trainer and 1–4 trainees per site. There was no statistically significant difference observed between trainer and trainee results (trainer and trainee mean = −0.028cm, 95% CI = −0.067 to −0.011, P = 0.1607; Figure 3)

Discussion

Our study contributes to the literature by demonstrating excellent intrarater and interrater reliability for ultrasound measurements to determine QMLT in healthy volunteers. These results reflect the ability to standardize training and the feasibility of this procedure in a range of trainee's including dietitians, nurses, and physiotherapists with no prior ultrasound experience. A sample of "normal" values is now available to compare measures from a study population.

Current concerns regarding this measurement tool are that muscle wasting may occur without changes in QMLT, and that water content may affect measurements.[3] Future research will aim to address these questions, as it may be that edema may not influence measurements with maximum ultrasound compression.

One limitation of our study was that not all trainers performed the measurements twice on each of the ten healthy volunteers, thus decreasing the strength of the study. The variation in trainees with unknown prior exposure to ultrasound may help explain the wide variation in ICC. Notably, the more practice a trainee received the higher the ICC, which likely also contributed to a wide variation in ICC. Furthermore, "normal" values obtained will be limited in their generalizability given the much younger age and BMI of healthy subjects, which may be reflective of physical activity, compared with the potential ICU population.

Given that excellent intrarater and interrater reliability for ultrasound measurements are attainable, and that QMLT is reflective of overall muscle mass, the next step is to apply this methodology to determine overall muscle mass in ICU patients. The next step is to validate the tool in comparison to magnetic resonance imaging (MRI), the gold standard, as a surrogate measure of LBM both as a single measurement or serial measurements over time. Yet, given the challenging ICU environment, coupling ultrasound with computed tomography (CT) scanning, which is more accessible for the ICU patient than the gold standard MRI, would provide additional insight into the validity of this ultrasound technique. Once research establishes bedside ultrasound measurements of QMLT as reliable and valid, providers will be able to assess muscle wasting longitudinally and measure effectiveness of nutrition or physical interventions meant to slow or reverse ICU-acquired muscle loss. US of the QMLT may ultimately be used to screen patients at risk for ICU-acquired muscle wasting upon admission and during hospitalization, and lead to muscle wasting prevention and intervention to decrease patient morbidity, mortality, and ICU length of stay.

Acknowledgments

References

1. Herridge MS, Cheung AM, Tansey CM, et al. Canadian Critical Care Trials Group. One-year outcomes in survivors of the acute respiratory distress syndrome. N Engl J Med. 2003; 348(8):683–693. [PubMed: 12594312]

2. Griffiths RD, Hall JB. Exploring intensive care unit-acquired weakness. Preface. Crit Care Med. 2009; 37(suppl):S295. [PubMed: 20046112]

3. Puthucheary Z, Montgomery H, Moxham J, Harridge S, Hart N. Structure to function: muscle failure in critically ill patients. J Physiol. 2010; 588(pt 23):4641–4648. [PubMed: 20961998]

4. De Jonghe BM, Sharshar T, Lefaucheur JP, et al. Groupe de Reflexion et d'Etude des Neuromyopathies en Reanimation. Paresis acquired in the intensive care unit: a prospective multicenter study. JAMA. 2002; 288(22):2859–2867. [PubMed: 12472328]

5. Sharshar T, Bastuji-Garin S, Stevens RD, et al. Groupe de Reflexion et d'Etude des Neuromyopathies en Reanimation. Presence and severity of intensive care unit-acquired paresis at time of awakening are associated with increased intensive care unit and hospital mortality. Crit Care Med. 2009; 37(12):3047–3053. [PubMed: 19770751]

6. Ali NA, O'Brien JM Jr, Hoffmann SP, et al. Midwest Critical Care Consortium. Acquired weakness, handgrip strength, and mortality in critically ill patients. Am J Respir Crit Care Med. 2008; 178(3):261–268. [PubMed: 18511703]

7. Hermans G, Clearckx B, Vanhullebusch T, et al. Interobserver agreement of Medical Research Council sum-score and handgrip strength in the intensive care unit. Muscle Nerve. 2012; 45(1):18–25. [PubMed: 22190301]

8. Thomaes T, Thomis M, Onkelinx S, Coudyzer W, Cornelissen V, Vanhees L. Reliability and validity of the ultrasound technique to measure the rectus femoris muscle diameter in older CAD-patients. BMC Med Imaging. 2012; 12:7.10.1186/1471-2342-12-7 [PubMed: 22471726]

9. Mourtzakis M, Prado CM, 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(5):997–1006.10.1139/H08-075 [PubMed: 18923576]

10. Arbeille P, Kerbeci P, Capri A, Dannaud C, Trappe SW, Trappe TA. Quantification of muscle volume by echography: comparison with MRI data on subjects in long-term bed rest. Ultrasound Med Biol. 2009; 35(7):1092–1097. [PubMed: 19394755]

11. Campbell IT, Watt T, Withers D, et al. Muscle thickness, measured with ultrasound, may be an indicator of lean tissue wasting in multiple organ failure in the presence of edema. Am J Clin Nutr. 1995; 62(3):533–539. [PubMed: 7661114]

12. Gruther W, Benesch T, Zorn C, et al. Muscle wasting in intensive care patients: ultrasound observation of the m. quadriceps femoris muscle layer. J Rehabil Med. 2008; 40(3):185–189. [PubMed: 18292919]
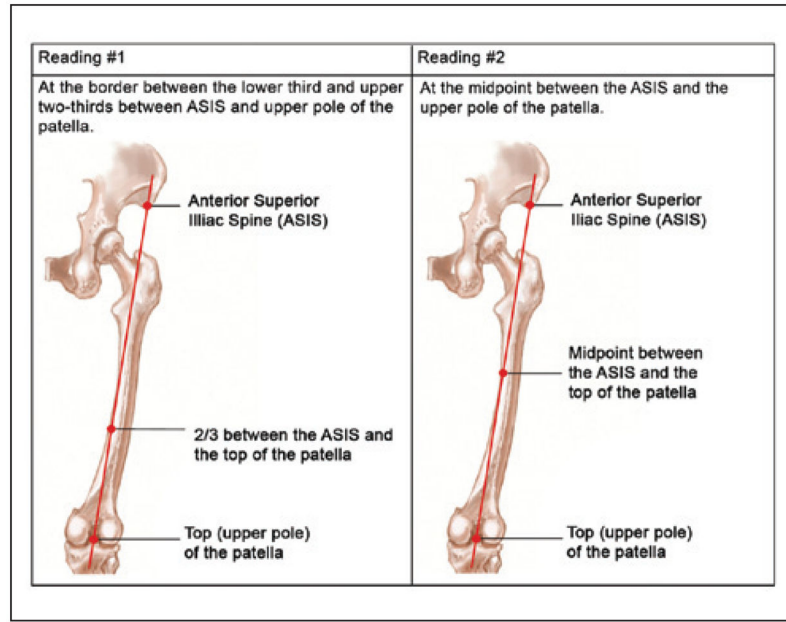
Figure 1.
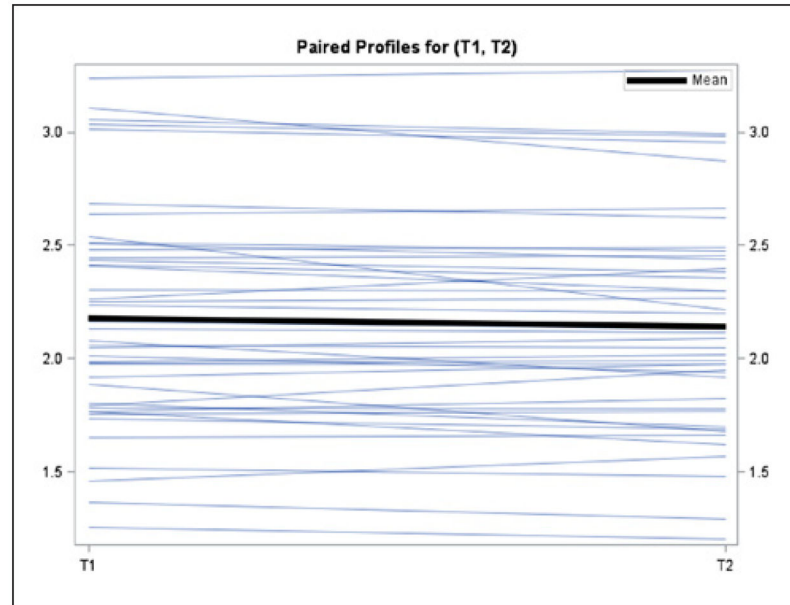Quadriceps muscle layer thickness measurements.

Figure 2.
Mean difference between trainer measurement 1 (T1) and 2 (T2). Mean (95% CI) =0.033 (0.004 to 0.061) p=0.0247; Y-axis (cm).
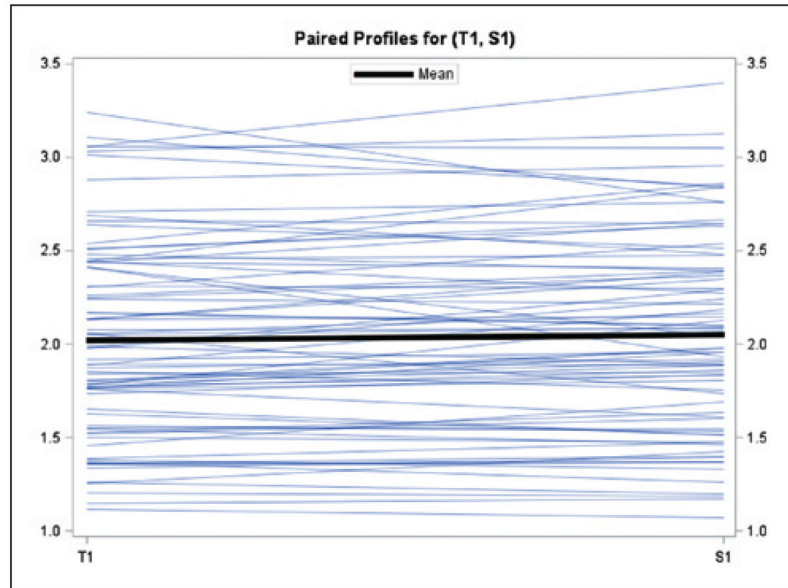
Figure 3.
Mean difference between trainer (T1) and trainee (S1). Mean (95% CI) =−0.0276 (−0.0665 to 0.0112), p=0.1607; Y-axis (cm).

Table 1

Demographics of Study Participants.

Variable	N	Mean	SD	Min	Max
Age (yr)	64[a]	30.6	8.4	21.0	55.0
Male	28[a]				
Female	40[a]				
Height (cm)	78	171.4	11.5	135.0	196.0
Weight (kg)	78	71.2	16.4	46.0	136.4
BMI	78	24.1	4.4	16.9	40.7

BMI, body mass index.

[a] Age was missing for 14 subjects and sex was missing from 10 subjects.

Table 2

Reliability Within Trainer (ie, intrarater reliability).

Site	Subjects Used	Between-Subject Variance	Within-Subject Variance	ICC
1	10	0.2307	0.01380	0.94
2	0	NA	NA	NA
3	10	0.2425	0.001018	>0.99
4	5	0.2819	0.000864	>0.99
5	12	0.05866	0.003180	0.95
6	5	0.2869	0.000613	>0.99
7	0	NA	NA	NA
	42	0.2642	0.004616	0.98

ICC, intraclass correlation coefficient; NA, not applicable.

Table 3

Reliability Between Trainer and Trainee for Measuring Same Subject (ie, interrater reliability).

Site	Subjects Used	Number of Different Trainees	Between-Subject Variance	Within-Subject Variance	ICC
1	10	3	0.2194	0.02900	0.88
2	13	2	0.3217	0.000769	>0.99
3	10	2	0.2305	0.02072	0.92
4	18	1	0.2623	0.009972	0.96
5	12	4	0.03587	0.01360	0.73
6	5	2	0.1714	0.02746	0.86
7	10	2	0.2030	0.01852	0.92
	78	16	0.2614	0.01503	0.95

ICC, intraclass correlation coefficient.