Published in final edited form as:

Muscle Nerve. 2014 July ; 50(1): 124–128. doi:10.1002/mus.24117.

MINIMAL TRAINING IS REQUIRED TO RELIABLY PERFORM QUANTITATIVE ULTRASOUND OF MUSCLE

CRAIG M. ZAIDMAN, MD¹, JIM S. WU, MD², SARAH WILDER, BS³, BASIL T. DARRAS, MD⁴,
and SEWARD B. RUTKOVE, MD³

¹Departments of Neurology and Pediatrics, Washington University School of Medicine, 660 South
Euclid Avenue, Box 8111, St. Louis, Missouri 63110, USA

²Department of Radiology, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA

³Department of Neurology, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA

⁴Department of Neurology, Boston Children's Hospital, Boston, Massachusetts, USA

Abstract

Introduction—Quantitative ultrasound can measure skeletal muscle pathology. We investigated whether inexperienced evaluators could accurately obtain and analyze ultrasound images.

Methods—Two examiners underwent a 20-minute training session before obtaining ultrasound images of several limb muscles in 21 healthy boys and 19 boys with Duchenne muscular dystrophy (DMD). Gray scale levels (GSLs) of muscle and subcutaneous fat were then measured by 2 analysts: a trained research assistant and a radiologist. We compared results between examiners and analysts.

Results—Interrater reliability of muscle GSLs was high between examiners (ICC = 0.85) and analysts (ICC = 0.84). As anticipated, GSLs were higher in dystrophic than in healthy muscles ($P < 0.001$). Fat GSLs were less reliable (ICC = 0.5–0.89) than muscle and increased with age and body size.

Conclusion—GSLs from ultrasound images of healthy and dystrophic skeletal muscle, but not from subcutaneous fat, can be obtained reliably and can be analyzed by inexperienced evaluators with minimal training.

Keywords

children; Duchenne muscular dystrophy; muscle; myopathy; quantitative ultrasound

Muscle ultrasound can be used to evaluate children with neuromuscular disorders and is a promising adjunct to the physical examination for quantitative assessment of patients with Duchenne muscular dystrophy (DMD), one of the most common, progressive, and devastating neuromuscular disorders of childhood.¹ The pathology of DMD, increased amounts of intramuscular fat and fibrosis, results in increasing reflection of ultrasonic

echoes that produce a brighter ultrasound image.^{2,3} The degree of brightness, or echogenicity, of dystrophic muscle relates to the severity and progression of the disease. In dystrophic muscle, echogenicity increases with age more in boys and young men with DMD than in the less severe Becker muscular dystrophy. In boys with DMD, the degree of echogenicity correlates with measures of strength and function and is similarly sensitive to changes in pathology over time.^{4,5} As ultrasound can be performed quickly and painlessly at the bedside and with minimal patient cooperation or effort, it could be used potentially to evaluate very young or very weak patients with DMD, in whom levels of strength and function are difficult to quantify.

Although the amount of brightness in the ultrasound image can be evaluated using either qualitative^{6,7} or quantitative^{2,8,9} techniques, quantification of the degree of echogenicity improves reliability and sensitivity to disease¹⁰ and is appealing as a potential outcome measure in clinical trials. One quantitative method to evaluate the ultrasound image is to measure the level of the gray scale pixels in the image using commonly available image analysis software, such as Adobe Photoshop or National Institutes of Health (NIH) ImageJ.^{2,11} When applied to skeletal muscle, quantitative ultrasound is sensitive to the presence of neuromuscular pathologies⁹ and can be measured reliably between different ultrasound systems by referencing images from muscle to a common phantom.^{12,13} An alternative method for standardizing gray scale level (GSL) measurements between ultrasound systems has been proposed that references the GSL from muscle to subcutaneous fat¹⁴; this approach would not require use of a common phantom, but it would require that GSLs measured from fat did not vary with disease or other subject characteristics.

Prior studies of quantitative muscle ultrasound have evaluated primarily images obtained by examiners with considerable ultrasound experience and training.^{4,8,10,14} Having skilled ultrasonographers and radiologists perform the examination and quantification, however, is problematic because it limits the applicability of the technique in the office and in multicenter clinical trials. Thus, in this study we evaluated whether inexperienced examiners could perform ultrasound imaging reliably after only a brief training session in boys with DMD and age-matched normal subjects. We also compared the reliability of image quantification between a board-certified musculoskeletal radiologist and a research assistant with only basic training.

METHODS

This protocol was approved by the institutional review board of Children's Hospital Boston. All subjects/guardians gave written consent/assent. We enrolled 19 boys, aged 2–13 [mean (SD): 7.7 (3.3)] years with a dystrophinopathy and a clinical presentation consistent with DMD, and 21 healthy control boys, aged 6 months to 12 [7.4 (3.3)] years. Controls had no history of weakness or neuromuscular disorders. Age, height, and weight were similar (P 0.2) between the DMD and control groups. Nine subjects with DMD were taking corticosteroids at the time of the ultrasound examination.

We obtained ultrasound images of the deltoid, elbow flexors (biceps brachii and brachialis), anterior forearm muscles, rectus femoris, tibialis anterior, and medial gastrocnemius on the

dominant side using a handheld ultrasound system (Terason 2000; Terason, Burlington, Massachusetts) with a linear 5–12-MHz probe. The depth, gain, compression, and time-gain compensation sonographic settings were kept constant between subjects, and all images were acquired on this single ultrasound device. Subjects were seated with the knee bent and the arm supported by a pillow at approximately mid-thoracic height and extended anteriorly at the shoulder with elbow extended and hand supinated and open. The probe was oriented transverse to the muscle length and perpendicular with the skin and muscle. The location for ultrasound probe placement for each muscle was marked on the limb at a specified proportion of the proximal-to-distal limb segment length at the midpoint of the muscle belly for each muscle studied.

A single jpeg image from each ultrasound examination of each muscle group was exported from the ultrasound system without any adjustments or manipulations, and the median GSL within a region of interest was measured using the histogram function on Adobe Photoshop (Adobe Systems, Inc., San Jose, California). A region of interest, including the entire depth of muscle between the subcutaneous tissue and deep bone or fascia (excluding the lateral margins of the imaged muscle), was selected using Photoshop, and the median GSL level was calculated. If the bone reflection was absent (due to severely abnormal signal in the overlying muscle) the region of interest was drawn within the muscle and superficial to the deep portion of the image, as identified by marked attenuation of the echo signal. Subcutaneous fat thickness in the ultrasound image was measured using calipers at each muscle site, and the average subcutaneous fat thickness for each subject was calculated. The GSL in a region of interest within the subcutaneous fat was also measured. All subjects with DMD and 20 control subjects had measurements of subcutaneous fat thickness and GSLs performed. Standing height and weight were recorded in most controls ($n = 19$). In ambulatory boys with DMD, standing height ($n = 14$) and weight ($n = 16$) were recorded when feasible.

Image acquisition was performed by examiners who were not healthcare professionals and who had no prior ultrasound experience except for a brief (20-minute) training session in the ultrasound protocol. First, a single examiner obtained images using ultrasound from each muscle group. Then a second examiner performed repeat image acquisition of the elbow flexors and rectus femoris after repositioning the patient. For intrarater reliability, the first examiner then repeated image acquisition of the elbow flexors. GSLs of the exported ultrasound images were measured independently by 2 image analysts. Image analyst 1 was an experienced musculoskeletal radiologist (J.W.), and image analyst 2 was a research assistant (S.W.) who lacked specific prior ultrasound experience, except for a brief training period during which she was instructed in region-of-interest placement and identifying the superficial fat, fascia, muscle, and bone. We compared results of images obtained by the 2 examiners (image acquisition) and measurements of the 2 image analysts (image quantification) separately. To determine the reliability of image acquisition, we compared GSLs, as measured by image analyst 1, of images obtained by the 2 examiners. To determine the reliability of image quantification, we compared gray scale measurements of image analysts 1 and 2 from ultrasound images obtained by examiner 1.

Statistics were performed using SPSS, version 14.0 (IBM SPSS, Inc., Armonk, New York). All values are expressed as mean (standard deviation) unless otherwise stated. An average score for each subject was determined by averaging results from each examined muscle group. Parametric distributions were confirmed using the Kolmogorov–Smirnov test. Intraclass correlation coefficients (ICCs) were determined using the 2-way mixed model for absolute agreement of single measures. Systematic bias between image analysts was assessed using Bland–Altman plots and paired *t*-tests. Ultrasound results of images obtained by examiner 1 and quantified by image analyst 2 were evaluated with a *t*-test for group comparisons and with the Pearson correlation coefficient (*r*) to define relationships to subject characteristics.

RESULTS

Quantitative Ultrasound of Muscle GSLs

Reliability of image acquisition and quantification of muscle GSLs was high for all muscles examined (Table 1) and across a wide range of GSLs (Figs. 1 and 2). GSLs from images of the elbow flexors and rectus femoris obtained by the 2 examiners were highly reliable (ICC = 0.85). There was no difference ($P > 0.4$) in the GSLs from the images obtained by the 2 examiners from either the elbow flexors or rectus femoris. Correlation of GSLs of 2 images of the elbow flexors obtained by a single examiner (intrarater reliability) was also high (ICC = 0.93). GSLs were also quantified similarly between the 2 analysts. GSLs measured from a single image by the 2 different analysts were highly reliable (ICC = 0.84). Image analyst 1 showed only a small bias toward lower average muscle GSLs compared with image analyst 2 [difference between image analysts 1 and 2: -2.7 (3.2) GSL, $P < 0.001$; Fig. 2].

The average muscle GSL was higher in DMD boys [49.3 (8.4)] than in controls [26.0 (4.0)] ($P < 0.001$). The median GSL from each muscle group was also higher in DMD boys than in controls (all $P < 0.001$). DMD boys as young as age 2 years showed higher average muscle GSLs than same-age controls (Fig. 3). Average muscle GSL in DMD was similar in boys taking steroids compared with those who were not ($P = 0.3$); boys not taking steroids were, on average, younger [6.0 (2.6) years] than those taking steroids [9.4 (1.8) years] ($P = 0.03$). In DMD patients, average muscle GSL trended higher with increasing age ($r = 0.4$, $P = 0.1$; Fig. 3), but did not vary with height ($n = 14$), weight ($n = 16$), or subcutaneous fat thickness (all $r < 0.2$, $P > 0.5$). In controls, average muscle GSL did not vary with age, height, weight, or subcutaneous fat thickness (all $r < 0.2$, $P > 0.3$).

Quantitative Ultrasound of Subcutaneous Fat GSLs

Reliability of image acquisition and quantification of subcutaneous fat GSLs was generally lower than in GSLs from muscle (Table 1). GSLs of subcutaneous fat from images of the elbow flexors and rectus femoris regions obtained by the 2 examiners had good reliability (ICC = 0.71 and 0.74, respectively), but were less reliable than those of muscle. GSLs of subcutaneous fat measured from a single image by the 2 different analysts varied by body region (ICC = 0.5–0.89) and were all lower than in muscle. Image analyst 1 showed a small bias toward higher average subcutaneous fat GSLs compared with image analyst 2 [difference between image analyst 1 and 2: 1.8 (2.6) GSL, $P < 0.001$].

In controls, the average subcutaneous fat GSL was higher with increased age ($r = 0.6$, $P = 0.008$) and height ($r = 0.5$, $P = 0.03$), trended to increase with greater weight ($r = 0.4$, $P = 0.1$), was lower with increased average muscle GSL ($r = -0.5$, $P = 0.03$), and did not vary with subcutaneous fat tissue thickness ($r = -0.1$, $P = 0.7$). Compared with controls, boys with DMD had lower average subcutaneous fat GSLs in the arms but not in the legs. The subcutaneous fat GSL was lower in DMD than in controls from areas overlying the deltoid [26.1 (5.1) vs. 30.9 (6.1), $P = 0.01$], biceps brachii [33.1 (6.0) vs. 38.0 (6.3), $P = 0.03$], and anterior forearm muscles [31.5 (6.7) vs. 37.5 (7.0), $P = 0.02$]. The subcutaneous fat GSL was similar between boys with DMD and controls from sites overlying the quadriceps, tibialis anterior, and medial gastrocnemius (all $P > 0.4$).

DISCUSSION

Reliable measures of quantitative ultrasound of skeletal muscle can be obtained by examiners without prior healthcare or sonography experience after a limited training session. Quantitative ultrasound of muscle includes 2 steps: the images must be obtained and then quantified. We found that 2 examiners, each with only a brief training session, were able to obtain images of the elbow flexors and rectus femoris, 2 commonly imaged muscles evaluated in DMD, which yielded highly reliable GSLs (ICC = 0.85) when measured by a single image analyst. Additional studies are required to determine the reliability of examiners for obtaining images of other muscles. In addition, we found that when evaluating the same ultrasound image, an image analyst with no prior ultrasound experience except for a brief training session performed similarly (ICC = 0.84) to an experienced musculoskeletal radiologist. These findings are comparable to those of the only other single study showing greater reliability of quantitative than qualitative analysis of ultrasound images of muscle analyzed by an experienced and inexperienced examiner.¹⁰

The images obtained in our study from trained examiners without extensive ultrasound experience were clinically relevant and similar to those obtained in earlier studies with experienced sonographers.^{5,8,9,13} As in prior studies, we found that muscle GSLs are higher in boys with DMD than in controls and trended higher in DMD boys, but not controls, with increasing age. An effect of age may have confounded our ability to determine whether steroids impact the GSL of muscle, as the boys in our cross-sectional sample who were not taking steroids were, on average, 3 years younger than those on treatment. Larger longitudinal studies are underway comparing changes in ultrasound to assessments of strength and function over time and after treatment.

One potential limitation of GSL analysis of ultrasound images is that results are highly dependent on the ultrasound system and settings. Reliability between ultrasound systems can be improved by standardizing the ultrasound settings and by referencing images to a shared phantom.^{12,15} An alternative approach has been proposed to reference results from muscle to subcutaneous fat.¹⁴ However, in controls, the GSL of fat varied with age and body size. GSLs of subcutaneous fat were also higher in arm regions of boys in the control group than in boys with DMD. Although GSLs measured from subcutaneous fat showed generally good reliability, this reliability varied depending on body region. For instance, reliability of GSLs from subcutaneous fat was particularly low when measured over the biceps brachii (ICC =

0.5). This may have been because the subcutaneous fat layer in this region is very thin, thus making reliable measurements more difficult to obtain. One study also found that referencing GSLs of muscle to subcutaneous fat does not control for variations in different ultrasound configurations.¹⁵ These findings suggest that subcutaneous fat should not be used as a reference value for standardizing ultrasound echogenicity measurements and that reference to a common phantom should be performed instead to standardize measurements between different systems.

We identified a small systemic bias between image analysts (2.5 GSLs). This is unlikely to have a large clinical effect, as this bias is small compared with the average difference between dystrophic and healthy muscle (23.3 GSLs). However, measurement bias could obscure very small effects, such as those that may be expected when evaluating repeated measures over time, and can be reduced by using the same image analyst to perform the measurements, regardless of level of training and expertise. One potential explanation for variations in GSL measurement between analysts is differences in identifying muscle borders for standardized region-of-interest placement. In particular, the deep border of the muscle can sometimes be difficult to visualize when muscles are very large, deep, or echogenic. One solution is to quantify only the superficial region within the muscle. Prior studies that measured GSLs from only the most superficial region within muscle detected disease progression over time in DMD⁴ and showed improved reliability between ultrasound systems compared with analyses of larger regions of muscle.¹³

In conclusion, reliable and clinically valid images of muscle may be obtained using ultrasound and can be quantified by trained non-professional evaluators without prior experience. This finding will simplify the implementation of quantitative ultrasound of skeletal muscle in multicenter clinical therapeutic trials and for regular day-to-day care of neuromuscular disease patients.

Acknowledgments

The study was supported by the National Institutes of Health (R01 AR060850).

The authors thank Rebecca Parad and Elizabeth Shriber for assistance with data collection and study coordination.

Abbreviations

DMD	Duchenne muscular dystrophy
GSL	gray scale level

References

1. Jeppesen J, Green A, Steffensen BF, Rahbek J. The Duchenne muscular dystrophy population in Denmark, 1977–2001: prevalence, incidence and survival in relation to the introduction of ventilator use. *Neuromuscul Disord.* 2003; 13:804–812. [PubMed: 14678803]
2. Heckmatt J, Rodillo E, Doherty M, Willson K, Leeman S. Quantitative sonography of muscle. *J Child Neurol.* 1989; 4(suppl):S101–106. [PubMed: 2681371]

3. Pillen S, Tak RO, Zwarts MJ, Lammens MM, Verrijp KN, Arts IM, et al. Skeletal muscle ultrasound: correlation between fibrous tissue and echo intensity. *Ultrasound Med Biol*. 2009; 35:443–446. [PubMed: 19081667]
4. Jansen M, van Alfen N, Nijhuis van der Sanden MW, van Dijk JP, Pillen S, de Groot IJ. Quantitative muscle ultrasound is a promising longitudinal follow-up tool in Duchenne muscular dystrophy. *Neuromuscul Disord*. 2012; 22:306–317. [PubMed: 22133654]
5. Zaidman CM, Connolly AM, Malkus EC, Florence JM, Pestronk A. Quantitative ultrasound using backscatter analysis in Duchenne and Becker muscular dystrophy. *Neuromuscul Disord*. 2010; 20:805–809. [PubMed: 20817454]
6. Heckmatt JZ, Dubowitz V, Leeman S. Detection of pathological change in dystrophic muscle with B-scan ultrasound imaging. *Lancet*. 1980; 1:1389–1390. [PubMed: 6104175]
7. Brockmann K, Becker P, Schreiber G, Neubert K, Brunner E, Bonnemann C. Sensitivity and specificity of qualitative muscle ultrasound in assessment of suspected neuromuscular disease in childhood. *Neuromuscul Disord*. 2007; 17:517–523. [PubMed: 17537637]
8. Zaidman CM, Holland MR, Anderson CC, Pestronk A. Calibrated quantitative ultrasound imaging of skeletal muscle using backscatter analysis. *Muscle Nerve*. 2008; 38:893–898. [PubMed: 18563722]
9. Pillen S, Verrips A, van Alfen N, Arts IM, Sie LT, Zwarts MJ. Quantitative skeletal muscle ultrasound: diagnostic value in childhood neuromuscular disease. *Neuromuscul Disord*. 2007; 17:509–516. [PubMed: 17537635]
10. Pillen S, van Keimpema M, Nievelstein RA, Verrips A, van Kruijsbergen-Raijmann W, et al. Skeletal muscle ultrasonography: visual versus quantitative evaluation. *Ultrasound Med Biol*. 2006; 32:1315–1321. [PubMed: 16965971]
11. Scholten RR, Pillen S, Verrips A, Zwarts MJ. Quantitative ultrasonography of skeletal muscles in children: normal values. *Muscle Nerve*. 2003; 27:693–698. [PubMed: 12766980]
12. Pillen S, van Dijk JP, Weijers G, Raijmann W, de Korte CL, Zwarts MJ. Quantitative gray-scale analysis in skeletal muscle ultrasound: a comparison study of two ultrasound devices. *Muscle Nerve*. 2009; 39:781–786. [PubMed: 19301363]
13. Zaidman CM, Holland MR, Hughes MS. Quantitative ultrasound of skeletal muscle: reliable measurements of calibrated muscle backscatter from different ultrasound systems. *Ultrasound Med Biol*. 2012; 38:1618–1625. [PubMed: 22763008]
14. Wu JS, Darras BT, Rutkove SB. Assessing spinal muscular atrophy with quantitative ultrasound. *Neurology*. 2010; 75:526–531. [PubMed: 20697104]
15. Hobson-Webb LD, Mhoon JT, Juel VC. Effect of transducer frequency on muscle luminosity ratio. *Muscle Nerve*. 2011; 44:612–613. [PubMed: 21922478]

FIGURE 2.

Bland–Altman plots of gray scale levels of muscle measured by 2 analysts. The average muscle GSLs quantified by experienced (rater 1) and inexperienced (rater 2) evaluators are also similar across a wide range of GSLs and show only a small bias (mean difference -2.7 GSL) between image analysts.

FIGURE 3.

Average muscle gray scale levels by age in healthy boys and boys with DMD. Average muscle GSL is higher in DMD than in controls ($P < 0.001$) even in children as young as 2 years, and there was a trend to increase with age in boys with DMD ($r = 0.4$, $P = 0.1$) but not in healthy controls ($r = -0.2$, $P = 0.5$).

Table 1

Interrater reliability (ICC) for GSLs measured from muscle and from subcutaneous fat.

Comparison	Region	Muscle GSL (ICC)	Subcutaneous fat GSL (ICC)
Repeat image acquisition	Elbow flexors	0.88	0.71
	Rectus femoris	0.85	0.74
Repeat quantification	Deltoid	0.84	0.70
	Elbow flexors	0.93	0.50
	Finger flexors	0.95	0.89
	Rectus femoris	0.88	0.72
	Tibialis anterior	0.87	0.80
	Medial gastrocnemius	0.92	0.71

GSL, gray scale level; ICC, intraclass correlation coefficient.