Published in final edited form as:

J Orthop Sports Phys Ther. 2014 June ; 44(6): 425–434. doi:10.2519/jospt.2014.4584.

Ultrasound Imaging: Intra- and Inter-Examiner Reliability for Multifidus Muscle Thickness Assessment in Adults Aged 60–85 Years versus Younger Adults

J. Megan Sions, DPT, PhD¹, Teonette O. Velasco, DPT¹, Deydre S. Teyhen, PT, PhD², and Gregory E. Hicks, PT, PhD¹

¹Department of Physical Therapy, University of Delaware, Newark, DE, USA

²Telemedicine and Advanced Technology Research Center, U.S. Army Medical Research and Materiel Command, Ft Detrick, MD, USA

Abstract

Study Design—Intra- and inter-examiner, within- and between-days, reliability study

Objectives—To compare the intra and inter-examiner procedural reliability of ultrasound imaging (USI) thickness measurements of the lumbar multifidus muscle at rest and during a contralateral lower limb lift (CLL) between older (60–85 years) and younger (18–40 years) adults.

Background—Among younger adults, USI has been shown to be reliable and valid for assessing multifidi thickness. Older adults present with age-related changes that may impact assessment. To our knowledge, no USI studies have focused on establishing reliability for multifidi thickness assessments in older adults.

Methods—Two examiners performed assessments of lumbar multifidi thickness at rest and during a CLL in 30 older and 31 younger adults. Intraclass correlation coefficients (ICCs) with 95% confidence intervals (CI), standard error of measurement (SEMs), and 95% minimal detectable change (MDCs) for resting, CLL, and percent thickness change were calculated.

Results—Within-day, inter-examiner procedural reliability for multifidus thickness measurements were similar amongst younger (ICCs=.90–.92) and older adults (ICCs=.86–.90) as was, between-days intra-examiner and inter-examiner reliability (younger ICCs=.84–.94; older ICCs=.86–.93). Throughout, estimates of percent thickness change were inconsistent (ICCs=.31–.86) and SEMs and MDCs were larger for older adults.

Conclusions—USI reliability for measurements of multifidi thickness at rest and during a CLL amongst older adults is consistent with previously published findings for younger adults. Among older adults, it may be preferable to have a single examiner conduct USI assessments.

Please send correspondences to: J. Megan Sions at University of Delaware, 540 South College Ave., Suite 144A, Newark, DE 19713, Phone: 302-831-7142; Fax: 302-831-4468 or megsions@udel.edu.

This project was approved by the Institutional Review Board for Human Subjects Research at the University of Delaware, Newark, DE, USA.

The view(s) expressed herein are those of the author(s) and do not reflect the official policy or position of the U.S. Army Medical Research and Materiel Command, the U.S. Army Medical Department, the U.S. Army Office of the Surgeon General, the Department of the Army, Department of Defense, or the U.S. Government.

Keywords

aging/geriatrics; lumbar spine; multifidus; ultrasonography

Trunk-specific clinical measures, such as extensor strength and endurance are associated with balance and mobility performance in older adults.^{8, 23, 24} The multifidi muscles have been purported to be important in spinal stabilization,¹⁵ and thus, may play a key role in balance and mobility. While strength and endurance allow for gross assessment of posterior trunk muscles, ultrasound imaging (USI) may allow physical therapists to perform in-vivo assessments of multifidus muscle size and activity.^{10, 26} Intramuscular electromyography (EMG) is often considered the method of choice for evaluating multifidi activity,^{10, 20} but intramuscular EMG is generally precluded in point-of-care practice due to its complexity and invasiveness.²⁷ Prior studies have established reliability^{14, 28, 31, 36} and criterion-validity^{10, 27} for USI assessment of change in multifidus muscle thickness as a measure of multifidus activity in younger, healthy adults. Pending established USI reliability and validity for multifidi assessments in older adults, future investigations may explore relationships among USI assessments, trunk-specific clinical measures, balance, and physical performance.

As shown in TABLE 1, amongst younger adults, USI has been shown to be reliable for measuring multifidi thickness at rest and during various tasks (e.g. arm lift, leg lift).^{11, 25, 28, 33} Standard USI assessment techniques for the multifidi rely on the ability to delineate fascial lines from adjacent muscles.²¹ When viewing an ultrasound image, young muscle, with high fluid content, appears hypoechoic (dark with few shades of gray), while fascia, which has high collagen content, appears hyperechoic (white).³² Contrasting echogenicities of muscle and fascia enable tissue differentiation. With aging, there is an increase in fibrous tissue in muscles.¹⁴ Therefore, among older adults, echogenicities of muscle and fascia may be similar,²² resulting in more difficulty with differentiation of muscle from fascia.

Prior studies have established USI measurement reliability, particularly in younger adults.^{7, 10} *USI measurement reliability* may be defined as the ability of examiners to consistently perform measurements from previously obtained images. Establishing inter-examiner measurement reliability may allow clinicians and researchers to delegate measurements to trained personnel. *USI procedural reliability* may be defined as the ability of examiners to consistently perform the entire imaging process including measurements. Given that in clinical practice evaluations and re-evaluations may be conducted by different examiners, establishing inter-examiner, between-days procedural reliability may allow for real-world application of USI. As noted in TABLE 1, USI procedural reliability studies for assessment of multifidi thickness have focused on younger populations, with and without low back pain (LBP), but have not included individuals over 60 years.^{11, 25, 28, 33}

Traditionally, multifidus thickness, obtained via parasagittal imaging (FIGURE 1), is measured from the most dorsal aspect of the facet joint to the last dark pixel before the muscle's fascial line (FIGURE 2);²⁸ however, among older adults the fascial line may be indistinct from the muscle (FIGURE 3). An alternative measurement that includes the fascial

line may be necessary (FIGURE 3). However, inclusion of the fascial layer in the thickness measurement may not be ideal as inter-individuals comparisons may be limited due to variability in fascial thickness between individuals. With that being said, pre- and post-treatment comparisons within a single individual may have clinical utility, pending one accepts that the measure does not allow discrimination of changes in muscle and fascia thickness. Reliability comparisons between the standard and the proposed alternative technique may determine if the alternative approach is a viable secondary option.

To our knowledge, no studies have focused on establishing USI reliability for assessment of multifidi thickness among older adults. The primary objectives of this study were to (1) evaluate the intra- and inter-examiner reliability of USI for the assessment of multifidi thicknesses in older adults as compared to younger adults and (2) determine measurement errors and minimal detectable change (MDC) values for USI-obtained multifidi thicknesses in older adults. We hypothesized that measurement and procedural reliability would be good-to-excellent in older adults using a single examiner or 2 examiners but that USI reliability would be lower among older adults when compared to younger adults. The secondary objective was to compare USI measurement techniques that are inclusive and exclusive of the fascial layer. We hypothesized that when evaluating multifidus muscle thickness in older adults, an alternative technique that included the fascial layer would demonstrate better reliability than the standard convention.

METHODS

Participants

Thirty community-dwelling older (aged 60 to 85 years) and 31 younger (aged 18 to 40 years) adults without LBP, or a history of LBP, were recruited for comparison. All participants were excluded if he/she had (1) a history of low back surgery; (2) received services for LBP within the last 6 months; (3) experienced a recent traumatic event; (4) severely impaired mobility status, as defined by the use of an assistive device greater than a cane; (5) a neurological disorder or a terminal illness; (6) a diagnosis of scoliosis; or (7) difficulty lying prone. This study's protocol was accepted by the Human Subjects Institutional Review Board at the University of Delaware, all participants signed an informed consent, and data were collected and maintained in accordance with the approved protocol. Participants completed a demographics questionnaire and underwent body anthropometric measurements prior to USI.

USI Procedures

Two licensed physical therapists, trained in USI via continuing education courses focused on multifidi assessment, performed independent assessments of the participants. Examiner 1 had 1.5 years of USI experience in a research setting while Examiner 2 was a novice (i.e., conducted 10 USI sessions with examiner 1 and performed 10 imaging sessions independently prior to the start of the study that included older adult participants). Both examiners had in-depth anatomy training that included cadaveric dissection.

Participants completed 2 USI sessions within 10 days, with random determination of examiner order within each session. Between USI sessions, participants were asked to forgo any exercises targeting their trunk. Participation was terminated for those (n=3) who responded “yes” to the question: “Since the last session, have you participated in exercises targeting your belly or back regions?”

Lumbar extension increases erector spinae thickness,³⁰ therefore participants were placed prone with no more than 5 degrees of extension as determined with an inclinometer placed at the L4/5 interspinous space. Examiners marked the skin following palpation and confirmed skin markings as described by Wallwork et al.²⁸ All markings were removed prior to the second examiner to ensure that each examiner independently established transducer position.²⁸ Examiners were not able to observe one another imaging.²⁸

All parasagittal images were obtained using B-mode on a Mylab 25 portable ultrasonography unit (Biosound Esaote Inc., Indianapolis, IN) at 6.6 MHz using the tissue enhancement imaging feature, which maximizes the signal-to-noise ratio and improves resolution. Gain was adjusted for each image to optimize fascial lines. A 3.5–7.0 MHz^{10, 22, 26} curvilinear transducer²² was placed parasagittally and angled medially (FIGURE 1) to attain 3 images bilaterally of L4/5 multifidus thickness at rest and during a prone, isometric, contralateral lower limb lift (CLL) of about 5cm using a split-screen function.¹⁶ Participants performed 2 practice trials per limb prior to image acquisition; verbal feedback was not given after practice trials. During CLLs, where the trunk extensor muscles exert an outward pressure on the transducer, a slight counter-pressure was applied to minimize transducer displacement. Images were independent of one another, meaning the transducer was removed from the skin after each image. Image acquisition order was counterbalanced to control for systematic order effect.

USI Data Processing

Images were analyzed on the USI machine, using MyLab software, after participants had left; examiners were masked to measurement outputs. Both examiners completed the following standard measurements: 1) resting multifidus muscle thickness and 2) multifidus muscle thickness during the limb lift, as well as alternative thickness measurements, which included multifidi fascial lines, for both conditions. Measurements were entered into spreadsheets by a third member of the research team. Percent thickness change was calculated as $((\text{multifidus thickness during limb lift} - \text{resting thickness}) / \text{resting thickness}) * 100$.¹⁰

Data Analysis

Statistical analyses were performed using PASW Statistics 18 (SPSS, Inc., Chicago, IL). Two-way, ICCs with 95% CIs were used to estimate intra-examiner (model 3,3) and inter-examiner (model 2,3) reliability¹⁹ between-days using the average of 3 measurements. Based on proposed ICC cut-offs by Fleiss, ICCs of greater than .75 were considered excellent; 0.40 to .75 were considered fair-to-good.⁶ To evaluate potential differences in reliability, for different age groups and measurement techniques, we compared ICCs and corresponding 95% CIs for younger and older adults, as well as for the standard and

alternative USI measurement techniques. Standard error of measurement values (SEMs) and MDCs were calculated.^{5, 17} MDC was calculated as $1.96 * 2 * SEM$, where 1.96 derives from the 95% confidence interval of no change and the 2 is included because 2 measures were involved in measuring change.²

RESULTS

A total of 31 younger adults and 30 older adults participated. Demographics and anthropometrics are provided (TABLE 2). Older adults had higher body mass indices (BMIs) than younger adults ($P=.002$). There were no statistically significant differences between the groups for sex, race, or hand dominance. Of the 61 participants, 30 younger adults were eligible and 28 older adults were available for the second imaging session.

Measurement Reliability

Reliability is reported for the left side. Inter-examiner, measurement reliability point estimates were excellent for both groups for all measurements except percent thickness changes (TABLE 3). Percent thickness changes for younger adults (ICCs=.73, .74) were less reliable than for older adults (ICCs= .91, .93). In both groups, the alternative multifidus thickness measurement technique resulted in similar ICCs when compared to the standard technique. When older adults were compared to younger adults, SEMs and MDCs for multifidus resting thickness and multifidus thickness during a CLL were larger for older adults, but percent thickness change SEMs and MDCs were larger for younger adults.

Procedural Reliability

Within-day, procedural reliability, between examiners generally indicated higher reliability point estimates for younger adults as compared to older adults, with the exception of multifidi percent thickness changes (TABLE 4); however, overlapping 95% CIs indicate similar reliability. Alternative multifidi thickness measurements had similar reliability to standard measurements in both groups. With the exception of percent thickness change, when compared to older adults, measurements in younger adults had smaller SEMs and MDCs throughout.

Younger and older adults had similar between-days, inter-examiner procedural reliability for standard and alternative multifidi resting thicknesses and multifidi thicknesses during CLLs (TABLE 5). Between-days, inter-examiner procedural reliability had lower reliability point estimates (TABLE 5) than within-day procedural reliability (TABLE 4) for standard multifidi resting thickness and multifidus thickness during a CLL, but overlapping confidence intervals indicated similar reliability; SEMs and MDCs were greater for between-days reliability. In the older adult, the alternative measurement technique for multifidus thickness did not appear superior to the previously published standard technique per ICCs, SEMs, and MDCs (TABLE 5). Regarding between-days, intra-examiner reliability, overall examiner 1 had better reliability among older adults, while examiner 2 had better reliability among younger adults (TABLE 6).

DISCUSSION

Despite the theoretical challenges of using USI to measure muscle size in older adults, measurements of lumbar multifidi thickness in older adults, at rest and with a CLL, were found to have excellent reliability. While reliability was similar among older and younger adults for multifidi thickness assessments, comparisons of MDCs indicate greater changes pre-to-post treatment may be necessary to detect “true changes” in multifidi thicknesses among older adults. Use of the alternative measurement technique for assessing multifidi muscle thicknesses does not have improved reliability when compared to the standard technique, so use of the standard technique^{10, 12, 25, 29, 33} is recommended, in older adults.

There have been several USI reliability studies of multifidi thickness in younger adults.^{10, 11, 25, 28, 33} Our inter-examiner measurement reliability results for L4/5 multifidus thickness were similar to results previously published by Kiesel and colleagues (ICC=.95).¹⁰ Our ICCs may be slightly higher due to calculation of multifidus thickness from the average of 3 measurements, as opposed to a single measurement.¹⁰ Similar to findings of Wallwork and colleagues (TABLE 1),²⁸ our within-day, inter-examiner, procedural reliability was excellent for L4/5 multifidus thickness at rest in younger adults. Our between-days, inter-examiner, procedural reliability findings in both younger and older adults without LBP were similar to those of Koppenhaver and colleagues¹¹ in younger adults with LBP for L4/5 multifidus thickness at rest (ICC=.88) and multifidus thickness during a contralateral upper limb lift (ICC=.92). While we might expect ICCs to be lower for individuals with LBP in the Koppenhaver et al¹¹ study as compared to our healthy individuals, this was not the case. Greater measurement variability between-days, as indicated by larger SEMs for between-days analyses as compared to within-day analyses, may mask differences in ICCs between groups of individuals with and without pain. Comparison of our between-day, inter-examiner procedural reliability results to those of Koppenhaver et al¹¹ reveal smaller MDCs in younger adults without LBP and larger MDCs among older adults, when compared to younger adults with LBP. For our younger adults, mean multifidi resting thickness and mean multifidi thickness during a CLL were similar to those of previous studies in younger adults without LBP;^{25, 28, 33} however, mean thicknesses of our older adults were more akin to younger adults with LBP.¹¹

Differences in methodology are a consideration when comparing these findings to other USI studies. Among younger adults, typically the contralateral upper limb is used to assess multifidus thickness during a limb lifting task.^{10-12, 33} During our pilot testing, we found that many older adults had difficulty performing an upper limb lift; difficulty may be secondary to concurrent musculoskeletal issues, such as rotator cuff tears, adhesive capsulitis, and/or increased thoracic kyphosis. Seeking out alternates, we adopted the prone CLL used by Pressler and colleagues.¹⁶ In both our younger and older adults, this technique demonstrated excellent reliability.

Multifidi thicknesses have been used to assess multifidi function, where percent thickness change from a resting- to a contracted-state has been correlated to electromyographic activity at low level contractions.¹⁰ Recommendations are that ICCs, SEMs, and MDCs be calculated not just for multifidus resting thickness and thickness during a limb lift but also

for percent thickness change.³ Some may argue that if percent thickness change is unreliable then the clinical value of multifidi thickness measurements is questionable. However, for younger adults, despite excellent ICCs (.98) for standard multifidus thickness measurements (TABLE 3), percent thickness change had a lower ICC point estimate (.74). The reliability of the independent measurements did not change; therefore, these data may illustrate magnification of the measurement error that occurs by combining the measurements. Further, upon examination of the SEMs and MDCs for percent thickness change in TABLE 3, we would expect percent thickness change SEMs and MDCs to follow the general trend of smaller values for younger adults when compared to older adults. In fact, percent thickness change SEMs and MDCs for younger adults were larger than SEMs and MDCs for older adults. Thus, we suggest that the clinical utility of USI multifidi thickness assessments not be based on percent thickness change ICCs, SEMs, and MDCs. Cautious interpretation of percent thickness change findings is supported by Whittaker and Stokes,³¹ who acknowledge that multiple factors, e.g. multifidus resting state, muscle extensibility, fiber orientation, adjacent longissimus contraction, limitations associated with use of a 2-dimensional imaging technique, and slight alterations of the transducer angle, may influence USI percent thickness change assessments.

With training, physical therapists may be able to reliably perform USI assessments of lumbar multifidus thickness in the clinical setting. For example, Examiner 2, the more novice examiner, was able to attain good-to-excellent intra-examiner reliability for all measurements in both age groups. However, it is important to note that USI training should extend beyond image acquisition and measurements, including USI education on basic principles, recognition of artifacts, parameter selection, participant screening, and prudent use,³² and in-depth anatomy education that includes cadaveric dissection of the region-of-interest. Further, specificity of training may be important. For example, examiner 1, who predominantly trained via imaging older individuals, had better reliability than examiner 2 among older adults. In contrast, examiner 2, who predominantly trained via imaging in younger individuals, had better reliability than examiner 1 in younger adults.

Study Limitations

Increased BMI has been correlated to increased muscle echogenicity, which may have added to the challenge of detecting fascial lines in our older adults.¹³ Given that it would be difficult to recruit older adults without a history of LBP, we recruited older adults who had not sought services for LBP in the past 6 months. LBP has been associated with increased fibrosis content¹ and thickening and/or disorganization of the fascial layer.¹³ As muscle recovery may not be spontaneous following an episode of LBP,^{4,9} residual LBP-related changes cannot be discounted and may have negatively impacted reliability when imaging older adults. Further, while the alternative thickness measurement technique could be used in healthy, older adults, we do not advocate use of the alternative technique among individuals with recurrent or chronic LBP, as the fascial layer is thickened and disorganized.¹³ We did not account for group differences in loading, but acknowledge that exercises that strain muscle beyond everyday activities, may affect collagen content and fascia alignment.¹⁸ Thus, differences in USI reliability between younger and older adults may be, in part, secondary to any of the abovementioned factors.

Consistent with previously published USI reliability studies,^{10–12, 25} we did not standardize our CLL height through use of an external apparatus. It is possible that percent thickness change reliability was affected by this methodological decision. For example, consider the Koppenhaver et al¹¹ study, which did not use an apparatus, to the Wong et al³³ study that used an external apparatus to standardize limb lift height. Intra-examiner, within-day ICCs were higher for percent thickness changes in the Wong et al³³ study. Some may note that all participants in the Koppenhaver et al¹¹ study had LBP, while only 50% of those in the Wong et al³³ study had LBP, postulating that ICCs would be lower in the presence of LBP. However, at least among younger adults, the presence of LBP seems to have little impact on within-day, inter-examiner USI reliability as ICCs have been shown to be similar between younger adults with¹¹ and without²⁵ LBP. Future research is needed to determine if LBP presence affects between-days USI reliability in younger adults and USI reliability in older adults.

CONCLUSION

In summary, within- and between-days USI by 1 or more examiners is reliable for assessment of multifidus muscle resting thickness and multifidus muscle thickness during a CLL in older adults and comparable to younger adults. ICCs calculated for percent thickness change may or may not indicate acceptable reliability despite excellent reliability for the component measurements (i.e. resting thickness and thickness during a limb lift). Between-group comparisons of SEMs and MDCs across-days indicate that thickness changes may be easier to detect in younger adults when compared to older adults. When necessary due to concurrent comorbidities, in older adults, CLLs may be reliably used in lieu of upper limb lifts. Measurements of multifidus thickness can be performed as previously reported among younger adults. Future USI studies in geriatric populations may consider these findings.

Acknowledgments

The work of Dr. Sions is supported in part by the Foundation for Physical Therapy PODS I/II, the University of Delaware's Office of Graduate Studies, R21 HD057274 (NICHD), and 1R01AG041202-01(NIA). The work of Teonette Velasco and Dr. Hicks is supported in part by R21 HD057274 (NICHD) and 1R01AG041202-01 (NIA).

References

1. Barbat-Artigas S, Rolland Y, Zamboni M, Aubertin-Leheudre M. How to assess functional status: a new muscle quality index. *J Nutr Health Aging*. 2012; 16:67–77. [PubMed: 22238004]
2. Beckerman H, Roebroek ME, Lankhorst GJ, Becher JG, Bezemer PD, Verbeek AL. Smallest real difference, a link between reproducibility and responsiveness. *Qual Life Res*. 2001; 10:571–578. [PubMed: 11822790]
3. Costa LO, Maher CG, Latimer J, Hodges PW, Shirley D. An investigation of the reproducibility of ultrasound measures of abdominal muscle activation in patients with chronic non-specific low back pain. *Eur Spine J*. 2009; 18:1059–1065. [PubMed: 19415347]
4. D'Hooge R, Cagnie B, Crombez G, Vanderstraeten G, Dolphens M, Danneels L. Increased intramuscular fatty infiltration without differences in lumbar muscle cross-sectional area during remission of unilateral recurrent low back pain. *Man Ther*. 2012; 17:584–588. [PubMed: 22784801]
5. Eliasziw M, Young SL, Woodbury MG, Fryday-Field K. Statistical methodology for the concurrent assessment of interrater and intrarater reliability: using goniometric measurements as an example. *Phys Ther*. 1994; 74:777–788. [PubMed: 8047565]

6. Fleiss, JL. Wiley classics library. New York: Wiley; 1999. The design and analysis of clinical experiments.
7. Hebert JJ, Koppenhaver SL, Parent EC, Fritz JM. A systematic review of the reliability of rehabilitative ultrasound imaging for the quantitative assessment of the abdominal and lumbar trunk muscles. *Spine (Phila Pa 1976)*. 2009; 34:E848–856. [PubMed: 19927091]
8. Hernandez ME, Goldberg A, Alexander NB. Decreased muscle strength relates to self-reported stooping, crouching, or kneeling difficulty in older adults. *Phys Ther*. 2010; 90:67–74. [PubMed: 19942678]
9. Hides JA, Richardson CA, Jull GA. Multifidus muscle recovery is not automatic after resolution of acute, first-episode low back pain. *Spine (Phila Pa 1976)*. 1996; 21:2763–2769. [PubMed: 8979323]
10. Kiesel KB, Uhl TL, Underwood FB, Rodd DW, Nitz AJ. Measurement of lumbar multifidus muscle contraction with rehabilitative ultrasound imaging. *Man Ther*. 2007; 12:161–166. [PubMed: 16973400]
11. Koppenhaver SL, Hebert JJ, Fritz JM, Parent EC, Teyhen DS, Magel JS. Reliability of rehabilitative ultrasound imaging of the transversus abdominis and lumbar multifidus muscles. *Arch Phys Med Rehabil*. 2009; 90:87–94. [PubMed: 19154834]
12. Koppenhaver SL, Parent EC, Teyhen DS, Hebert JJ, Fritz JM. The effect of averaging multiple trials on measurement error during ultrasound imaging of transversus abdominis and lumbar multifidus muscles in individuals with low back pain. *J Orthop Sports Phys Ther*. 2009; 39:604–611. [PubMed: 19648721]
13. Langevin HM, Stevens-Tuttle D, Fox JR, et al. Ultrasound evidence of altered lumbar connective tissue structure in human subjects with chronic low back pain. *BMC Musculoskelet Disord*. 2009; 10:151. [PubMed: 19958536]
14. Mann CJ, Perdiguer E, Kharraz Y, et al. Aberrant repair and fibrosis development in skeletal muscle. *Skeletal muscle*. 2011; 1:21. [PubMed: 21798099]
15. Moseley GL, Hodges PW, Gandevia SC. Deep and superficial fibers of the lumbar multifidus muscle are differentially active during voluntary arm movements. *Spine (Phila Pa 1976)*. 2002; 27:E29–36. [PubMed: 11805677]
16. Pressler JF, Heiss DG, Buford JA, Chidley JV. Between-day repeatability and symmetry of multifidus cross-sectional area measured using ultrasound imaging. *J Orthop Sports Phys Ther*. 2006; 36:10–18. [PubMed: 16494069]
17. Roebroeck ME, Harlaar J, Lankhorst GJ. The application of generalizability theory to reliability assessment: an illustration using isometric force measurements. *Phys Ther*. 1993; 73:386–395. discussion 396–401. [PubMed: 8497513]
18. Schleip R, Muller DG. Training principles for fascial connective tissues: scientific foundation and suggested practical applications. *Journal of bodywork and movement therapies*. 2013; 17:103–115. [PubMed: 23294691]
19. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin*. 1979; 86:420–428. [PubMed: 18839484]
20. Stokes IA, Henry SM, Single RM. Surface EMG electrodes do not accurately record from lumbar multifidus muscles. *Clinical biomechanics*. 2003; 18:9–13. [PubMed: 12527241]
21. Stokes M, Hides J, Elliott J, Kiesel K, Hodges P. Rehabilitative ultrasound imaging of the posterior paraspinal muscles. *J Orthop Sports Phys Ther*. 2007; 37:581–595. [PubMed: 17970405]
22. Stokes M, Rankin G, Newham DJ. Ultrasound imaging of lumbar multifidus muscle: normal reference ranges for measurements and practical guidance on the technique. *Man Ther*. 2005; 10:116–126. [PubMed: 15922232]
23. Suri P, Kiely DK, Leveille SG, Frontera WR, Bean JF. Increased trunk extension endurance is associated with meaningful improvement in balance among older adults with mobility problems. *Arch Phys Med Rehabil*. 2011; 92:1038–1043. [PubMed: 21636073]
24. Suri P, Kiely DK, Leveille SG, Frontera WR, Bean JF. Trunk muscle attributes are associated with balance and mobility in older adults: a pilot study. *PM R*. 2009; 1:916–924. [PubMed: 19854420]
25. Teyhen DS, George SZ, Dugan JL, Williamson J, Neilson BD, Childs JD. Inter-rater reliability of ultrasound imaging of the trunk musculature among novice raters. *Journal of ultrasound in*

- medicine: official journal of the American Institute of Ultrasound in Medicine. 2011; 30:347–356. [PubMed: 21357556]
26. Van K, Hides JA, Richardson CA. The use of real-time ultrasound imaging for biofeedback of lumbar multifidus muscle contraction in healthy subjects. *J Orthop Sports Phys Ther.* 2006; 36:920–925. [PubMed: 17193869]
27. Vasseljen O, Dahl HH, Mork PJ, Torp HG. Muscle activity onset in the lumbar multifidus muscle recorded simultaneously by ultrasound imaging and intramuscular electromyography. *Clinical biomechanics.* 2006; 21:905–913. [PubMed: 16822599]
28. Wallwork TL, Hides JA, Stanton WR. Intrarater and interrater reliability of assessment of lumbar multifidus muscle thickness using rehabilitative ultrasound imaging. *J Orthop Sports Phys Ther.* 2007; 37:608–612. [PubMed: 17970407]
29. Wallwork TL, Stanton WR, Freke M, Hides JA. The effect of chronic low back pain on size and contraction of the lumbar multifidus muscle. *Man Ther.* 2009; 14:496–500. [PubMed: 19027343]
30. Watanabe K, Miyamoto K, Masuda T, Shimizu K. Use of ultrasonography to evaluate thickness of the erector spinae muscle in maximum flexion and extension of the lumbar spine. *Spine (Phila Pa 1976).* 2004; 29:1472–1477. [PubMed: 15223941]
31. Whittaker JL, Stokes M. Ultrasound imaging and muscle function. *J Orthop Sports Phys Ther.* 2011; 41:572–580. [PubMed: 21654098]
32. Whittaker JL, Teyhen DS, Elliott JM, et al. Rehabilitative ultrasound imaging: understanding the technology and its applications. *J Orthop Sports Phys Ther.* 2007; 37:434–449. [PubMed: 17877280]
33. Wong AY, Parent EC, Kawchuk GN. Reliability of Two Ultrasonic Imaging Analysis Methods in Quantifying Lumbar Multifidus Thickness. *J Orthop Sports Phys Ther.* 2013; 43:251–262. [PubMed: 23221501]

KEY POINTS**Findings**

In older adults, ultrasound imaging (USI) measurements of multifidus muscle resting thickness and multifidus muscle thickness during a contralateral limb lift, may be performed as previously reported in the literature among younger adults with excellent within-and between-days reliability. Lower limb lifts may be used in lieu of upper limb lifts when necessary.

Implication

USI assessments of multifidi thicknesses during limb lifts allow for in-vivo assessments of muscle thickness. Given the proposed role of multifidi muscles in spinal stabilization, future research may explore relationships between USI multifidi muscle assessments and balance and mobility performance in older adults.

Caution

In older adults, USI assessments of multifidi thickness may be used to assess muscle change over time, but to detect the smallest changes, a single examiner should conduct assessments.

FIGURE 1.

Parasagittal imaging with the transducer placed just lateral to the spinous processes and angled medially to view the L4/5 facet joint.

FIGURE 2.

On the left are ultrasound images of the multifidus obtained via parasagittal imaging at rest (A) and during a straight leg raise (B) in a younger adult. On the right are anatomical diagrams showing the standard measurement technique for multifidus thickness, where (a) represents multifidus muscle. The standard measurement is taken from the facet joint to the last dark pixel before the muscle's fascial line.

FIGURE 3.

On the left are ultrasound images of the multifidus obtained via parasagittal imaging at rest (A) and during a straight leg raise (B) in an older adult. On the right are anatomical diagrams showing the proposed alternative measurement technique for multifidus thickness, where (a) represents multifidus muscle and its fascia. The alternative measurement is taken from the facet joint to the first dark pixel beyond the muscle's fascial line.

TABLE 1

Ultrasound Imaging Procedural Reliability for L4/5 Multifidus Thickness Assessment.

Study	Sample	Transducer Frequency, Type	Average of (Number of Images)	Intra/Inter-Examiner Reliability	Within/Between-Days Reliability	Multifidus Mean Thickness (SD)	ICC (95% CI)	SEM	MDC (95%)
Wallwork et al 2007 ²⁸ resting (cm)	n=10; mean age: 30.8+/-18.1 years; status: no LBP	5 MHz, curvilinear	3	Intra	Within	2.82 (.25); 2.86 (.26)	.88-.95 (.68-.99)	.06-.09	NR
			3	Inter	Within	NR	.97 (.87, .99)	.05	NR
Koppenhaver et al 2009 ¹¹ resting (cm)	n=30; mean age: 42.4+/-11.4 years; status: LBP	2-5 MHz, curvilinear	2	Intra	Within	3.46 (.62)	.97 (.94, .99)	.10	.28
			2	Intra	Between	3.44 (.62)	.98 (.95, .99)	.09	.25
			2	Inter	Within	3.32 (.60)	.88 (.63, .95)	.21	.58
			2	Inter	Between	3.33 (.60)	.88 (.60, .95)	.21	.58
			2	Intra	Within	3.79 (.65)	.99 (.98, 1.00)	.06	.16
			2	Intra	Between	3.82 (.66)	.97 (.94, .99)	.11	.31
during limb lift (cm)			2	Inter	Within	3.75 (.64)	.93 (.85, .97)	.17	.47
			2	Inter	Between	3.77 (.66)	.92 (.82, .97)	.18	.51
			2	Intra	Within	9.8 (8.3)	.78 (.53, .89)	4.0	11.0
			2	Intra	Between	11.2 (8.7)	.79 (.56, .90)	4.0	11.0
% thickness change			2	Inter	Within	13.4 (11.0)	.45 (-.09, .73)	8.1	22.6
			2	Inter	Between	13.9 (10.7)	.73 (.42, .88)	5.5	15.3
Teyhen et al 2011 ²⁵ resting (cm)	n=21; mean age: 21.5+/-4.4 years; status: no LBP	5 MHz, curvilinear	3	Inter	Within	2.83 (.46)	.87 (.68, .95)	.16	.46
			3	Inter	Within	3.50 (.56)	.94 (.86, .98)	.14	.38
during limb lift (cm)									
Wong et al 2012 ³³ resting (cm)	n=27; mean age: 29.6+/-10.0 years; status: no LBP (n=13), LBP (n=14)	2-5MHz, curvilinear	2	Intra	Within	2.61 (.39)	.99 (.99, 1.00)	.02	.07
			2	Intra	Between	2.59 (.40)	.93 (.85, .97)	.11	.30
during limb lift (cm)									
% thickness change			2	Intra	Within	3.28 (.44)	.99 (.99, 1.00)	.03	.09
			2	Intra	Between	3.30 (.44)	.97 (.92, .99)	.08	.22
			2	Intra	Within	26.6 (12.0)	.99 (.97, .99)	1.4	3.8
			2	Intra	Between	28.3 (10.8)	.80 (.61, .91)	4.8	13.2

Abbreviations: cm=centimeters; CI=confidence interval; ICC=intraclass correlation coefficient; LBP=low back pain; MHz=megahertz; MDC=minimal detectable change; NR=not reported; %=percent; SD=standard deviation; SEM=standard error of measurement.

TABLE 2

Participant Demographics and Anthropometrics.

Variables	Younger Adults (n=31)	Older Adults (n=30)	p-value
Age (years)	20.9 (19.8, 22.0)	71.8 (70.1, 73.4)	.000
BMI (kg/m ²)	24.0 (22.9, 25.2)	27.0 (25.2, 28.5)	.002
Female *	20 (64.5)	22 (73.3)	.457
Caucasian *	29 (93.5)	26 (86.7)	.175
Right hand dominant *	29 (93.5)	28 (93.3)	.513

Abbreviations: BMI=body mass index; CI=confidence interval; kg=kilograms; m=meters; %=percent; SD=standard deviation.

* Means and standard deviations are reported, except for denoted variables, which are reported as number of participants (percentage of sample).

TABLE 3
Inter-Examiner Measurement Reliability for L4/5 Multifidus Thickness Assessment

Younger Adults (n=31)	Examiner 1 mean thickness (SD)	Examiner 2 mean thickness (SD)	ICC 95% CI	SEM	MDC (95%)
resting*	2.97 (0.44)	2.92 (0.44)	.98 (.96, .99)	0.06	0.15
during limb lift*	3.73 (0.46)	3.69 (0.43)	.98 (.96, .99)	0.06	0.15
% thickness change [†]	20.3 (6.0)	27.3 (9.6)	.74 (-.18, .92)	3.9	10.9
resting ALT*	3.26 (0.50)	3.26 (0.48)	.99 (.98, .99)	0.04	0.10
during limb lift ALT*	4.08 (0.51)	4.11 (0.50)	.99 (.97, .99)	0.05	0.14
% thickness change ALT [†]	20.1 (5.1)	26.8 (7.9)	.73 (-.19, .92)	3.1	8.6
Older Adults (n=30)					
resting*	3.36 (0.80)	3.30 (0.77)	.98 (.97, .99)	0.08	0.22
during limb lift*	4.04 (0.84)	3.97 (0.82)	.98 (.97, .99)	0.10	0.28
% thickness change [†]	21.4 (9.0)	21.4 (8.6)	.93 (.86, .97)	2.3	6.5
resting ALT*	3.67 (0.79)	3.66 (0.77)	.98 (.97, .99)	0.09	0.25
during limb lift ALT*	4.39 (0.83)	4.34 (0.83)	.99 (.98, .99)	0.07	0.19
% thickness change ALT [†]	20.6 (7.5)	19.6 (8.6)	.91 (.81, .96)	2.5	6.9

Abbreviations: ALT=alternate measurement technique; CI=confidence interval; ICC=intraclass correlation coefficient; MDC=minimal detectable change; SD=standard deviation; SEM=standard error of measurement.

* Resting and limb lift means, standard deviations, SEMs, and MDCs are provided in centimeters.

[†] Percent thickness change was calculated ((thickness during limb lift-resting thickness)/resting thickness)*100.

TABLE 4
Within-Day, Inter-Examiner Procedural Reliability for L4/5 Multifidus Thickness Assessment.

Younger Adults (n=31)	Examiner 1 mean thickness (SD)	Examiner 2 mean thickness (SD)	ICC 95% CI	SEM	MDC (95%)
resting*	2.97 (0.44)	2.93 (0.47)	.90 (.80, .95)	0.14	0.40
during limb lift*	3.73 (0.46)	3.70 (0.47)	.90 (.79, .95)	0.15	0.41
% thickness change [†]	20.3 (6.0)	27.2 (10.1)	.47 (-.07, .74)	10.1	28.2
resting ALT*	3.26 (0.50)	3.28 (0.53)	.91 (.83, .96)	0.15	0.41
during limb lift ALT*	4.08 (0.51)	4.10 (0.53)	.92 (.84, .96)	0.14	0.40
% thickness change ALT [†]	20.1 (5.1)	26.1 (9.6)	.59 (.02, .81)	4.4	12.4
Older Adults (n=30)					
resting*	3.38 (0.80)	3.22 (0.71)	.88 (.74, .94)	0.26	0.73
during limb lift*	4.06 (0.83)	3.94 (0.72)	.90 (.79, .95)	0.24	0.68
% thickness change [†]	21.1 (9.0)	24.3 (12.0)	.66 (.31, .83)	5.8	16.1
resting ALT*	3.71 (0.81)	3.53 (0.73)	.86 (.71, .93)	0.28	0.78
during limb lift ALT*	4.43 (0.85)	4.33 (0.69)	.88 (.75, .94)	0.27	0.74
% thickness change ALT [†]	20.4 (7.4)	24.1 (11.5)	.63 (.25, .82)	5.4	15.1

Abbreviations: ALT=alternate measurement technique; CI=confidence interval; ICC=intraclass correlation coefficient; MDC=minimal detectable change; SD=standard deviation; SEM=standard error of measurement.

* Resting and limb lift means, standard deviations, SEMs, and MDCs are provided in centimeters.

[†] Percent thickness change was calculated ((thickness during limb lift-resting thickness)/resting thickness)*100.

TABLE 5
Between-Days, Inter-Examiner Procedural Reliability for L4/5 Multifidus Thickness Assessment.

Younger Adults (n=30)	Examiner 1 mean thickness (SD)	Examiner 2 mean thickness (SD)	ICC (95% CI)	SEM	MDC (95%)
resting*	2.99 (0.44)	2.87 (0.52)	.84 (.66, .92)	0.19	0.53
during limb lift*	3.75 (0.46)	3.67 (0.54)	.88 (.75, .94)	0.17	0.48
% thickness change [†]	20.2 (6.0)	28.7 (12.1)	.31 (-.21, .64)	7.4	20.7
resting ALT*	3.28 (0.50)	3.22 (0.60)	.90 (.80, .95)	0.17	0.47
during limb lift ALT*	4.10 (0.50)	4.08 (0.62)	.90 (.80, .95)	0.17	0.48
% thickness change ALT [†]	20.0 (5.1)	28.0 (10.0)	.59 (.02, .81)	5.1	14.1
Older Adults (n=28)					
resting*	3.34 (0.81)	3.19 (0.73)	.86 (.70, .93)	0.29	0.80
during limb lift*	4.00 (0.82)	3.86 (0.69)	.89 (.76, .94)	0.26	0.71
% thickness change [†]	21.1 (9.1)	22.7 (13.9)	.71 (.38, .86)	5.8	16.3
resting ALT*	3.65 (0.80)	3.57 (0.82)	.87 (.72, .94)	0.29	0.82
during limb lift ALT*	4.35 (0.82)	4.23 (0.78)	.89 (.77, .95)	0.26	0.73
% thickness change ALT [†]	20.3 (7.5)	20.3 (12.8)	.49 (-.27, .75)	8.5	23.7

Abbreviations: ALT=alternate measurement technique; CI=confidence interval; ICC=intraclass correlation coefficient; MDC=minimal detectable change; SD=standard deviation; SEM=standard error of measurement.

* Resting and limb lift means, standard deviations, SEMs, and MDCs are provided in centimeters.

[†] Percent thickness change was calculated ((thickness during limb lift-resting thickness)/resting thickness)*100.

TABLE 6
Between-Days, Intra-Examiner Procedural Reliability for L4/5 Multifidus Thickness Assessment.

	Examiner 1					Examiner 2				
	Day 1: mean thickness (SD)	Day 2: mean thickness (SD)	ICC (95% CI)	SEM	MDC	Day 1: mean thickness (SD)	Day 2: mean thickness (SD)	ICC (95% CI)	SEM	MDC (95%)
Younger Adults (n=30)										
resting*	2.99 (0.44)	2.91 (0.52)	.85 (.69, .93)	0.19	0.52	2.94 (0.48)	2.87 (0.52)	.94 (.87, .97)	0.12	0.34
during limb lift*	3.75 (0.46)	3.72 (0.50)	.90 (.81, .95)	0.14	0.40	3.72 (0.47)	3.67 (0.54)	.92 (.83, .96)	0.14	0.39
% thickness change [†]	20.2 (6.0)	29.5 (14.3)	.38 (-.15, .69)	8.6	24.0	27.5 (10.2)	28.7 (12.1)	.82 (.62, .91)	4.6	12.9
resting ALT*	3.28 (0.50)	3.21 (0.55)	.89 (.78, .95)	0.17	0.47	3.29 (0.53)	3.22 (0.60)	.94 (.88, .97)	0.13	0.36
during limb lift ALT*	4.10 (0.50)	4.08 (0.50)	.90 (.79, .95)	0.16	0.44	4.13 (0.52)	4.08 (0.62)	.91 (.82, .96)	0.17	0.46
% thickness change ALT [†]	20.0 (5.1)	28.7 (12.4)	.46 (-.11, .74)	6.6	18.3	26.4 (9.7)	28.0 (10.0)	.86 (.71, .93)	3.7	10.3
Older Adults (n=28)										
resting*	3.34 (0.81)	3.25 (0.68)	.92 (.83, .96)	0.21	0.59	3.15 (0.69)	3.19 (0.73)	.90 (.78, .95)	0.22	0.62
during limb lift*	4.00 (0.82)	3.93 (0.73)	.92 (.83, .96)	0.22	0.61	3.88 (0.70)	3.86 (0.69)	.92 (.82, .96)	0.20	0.55
% thickness change [†]	21.1 (9.1)	22.0 (10.9)	.78 (.52, .89)	4.6	13.0	24.7 (12.3)	22.7 (13.9)	.69 (.34, .86)	4.0	11.1
resting ALT*	3.65 (0.80)	3.66 (0.73)	.92 (.83, .96)	0.22	0.60	3.48 (0.71)	3.57 (0.82)	.89 (.76, .94)	0.25	0.70
during limb lift ALT*	4.35 (0.83)	4.40 (0.78)	.93 (.84, .96)	0.21	0.59	4.27 (0.68)	4.23 (0.78)	.91 (.82, .96)	0.21	0.58
% thickness change ALT [†]	20.3 (7.53)	21.3 (10.0)	.71 (.38, .86)	4.3	11.9	24.5 (11.8)	20.3 (12.8)	.72 (.40, .87)	6.3	17.5

Abbreviations: ALT=alternate measurement technique; CI=confidence interval; ICC=intraclass correlation coefficient; MDC=minimal detectable change; SD=standard deviation; SEM=standard error of measurement.

* Resting and limb lift means, standard deviations, SEMs, and MDCs are provided in centimeters.

[†] Percent thickness change was calculated ((thickness during limb lift-resting thickness)/resting thickness)*100.