

SHANE L. KOPPENHAVER, MPT¹ • ERIC C. PARENT, PT, PhD² • DEYDRE S. TEYHEN, PT, PhD³
JEFFREY J. HEBERT, DC⁴ • JULIE M. FRITZ, PT, PhD⁵

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

Ultrasound imaging is increasingly used in rehabilitation to evaluate muscle morphology and function in patients with neuromusculoskeletal disorders such as low back pain (LBP).²⁷ This application, called rehabilitative ultrasound imaging (RUSI), has, to date, mostly focused on the transversus abdominis (TrA) and the lumbar multifidus (LM) muscles, because dysfunction

of these muscles has been linked to LBP.^{8-12,32,33} Some aspects of the reliability of TrA and LM muscle thickness measurements using RUSI have been established. While most studies report high reliability for muscle thickness and cross-sectional area measurements (intraclass correlation coefficients [ICCs] of greater than 0.85), standard error of measurement (SEM) and minimal detectable change (MDC) are usually large in comparison to mean measurements.^{6,14,22,24,29-31}

A few researchers have reported improvements in the reliability and precision of RUSI measurements by using the mean of 3 thickness measurements instead of a single measure of the TrA and LM.^{24,31} While both studies found good reliability (ICCs, >0.85) and precision (TrA SEM, <0.45 mm; LM SEM, <1.3 mm) with single measures, MDC for the TrA was relatively large compared to its mean thickness. For example, using the MDC for resting thickness,²⁴ Teyhen et al²⁸ calculated a single measurement of

- **STUDY DESIGN:** Clinical measurement, reliability study.
- **OBJECTIVES:** To investigate the improvements in precision when averaging multiple measurements of percent change in muscle thickness of the transversus abdominis (TrA) and lumbar multifidus (LM) muscles.
- **BACKGROUND:** Although the reliability of TrA and LM muscle thickness measurements using rehabilitative ultrasound imaging (RUSI) is good, measurement error is often large relative to mean muscle thickness. Additionally, percent thickness change measures incorporate measurement error from both resting and contracted conditions.
- **METHODS:** Thirty volunteers with nonspecific low back pain participated. Thickness measurements of the TrA and LM muscles were obtained using RUSI at rest and during standardized tasks. Percent thickness change was calculated with the

formula $(\text{thickness}_{\text{contracted}} - \text{thickness}_{\text{rest}}) / \text{thickness}_{\text{rest}}$. Standard error of measurement (SEM) quantified precision when using 1 or a mean of 2 to 6 consecutive measurements.

- **RESULTS:** Compared to when using a single measurement, SEM of both the TrA and LM decreased by nearly 25% when using a mean of 2 measures, and by 50% when using the mean of 3 measures. Little precision was gained by averaging more than 3 measurements.
- **CONCLUSION:** When using RUSI to determine percent change in TrA and LM muscle thickness, intraexaminer measurement precision appears to be optimized by using an average of 3 consecutive measurements. *J Orthop Sport Phys Ther* 2009;39(8):604-611. doi:10.2519/jospt.2009.3088
- **KEY WORDS:** abdominal muscles, lumbar spine, reliability, ultrasonography

¹PhD candidate, University of Utah, College of Health, Salt Lake City, UT; Officer, US Army. ²Assistant Professor, Department of Physical Therapy, University of Alberta, Clinical Scientist, Glenrose Rehabilitation Hospital, Edmonton, Alberta, Canada. ³Associate Professor, US Army-Baylor University Doctoral Program in Physical Therapy, San Antonio, TX; Officer, US Army. ⁴PhD Candidate, University of Utah, College of Health, Salt Lake City, UT; Clinical Staff, Department of Neurosurgery, University of Utah, Salt Lake City, UT. ⁵Associate Professor, University of Utah, College of Health, Salt Lake City, UT; Clinical Outcomes Research Scientist, Intermountain Health Care, Salt Lake City, UT. The University of Utah and Intermountain Health Care's Institutional Review Boards approved the protocol for this study. All investigation conformed to the protocol and the ethical and humane principles of research. The opinions or assertions contained herein are the private views of the Authors and are not to be construed as official or as reflecting the views of the Departments of the Army or Defense. Address correspondence to Shane Koppenhaver, 1416 Downington Ave, Salt Lake City, UT 84105. E-mail: shanekoppenhaver@mac.com

resting TrA thickness would require a 41% change for one to be 95% confident that a change occurred. However, an average of 3 measurements resulted in a reduction in the MDC value such that one could be equally confident that a change occurred with only a 17% change in resting muscle thickness.^{24,28} This improvement in measurement precision is likely even more important when using percent change in muscle thickness ($[\text{thickness}_{\text{contracted}} - \text{thickness}_{\text{rest}}] / \text{thickness}_{\text{rest}}$), as these more clinically relevant measures incorporate the measurement error from both resting and contracted conditions.

While it is intuitive that the mean of multiple measures would reduce inter-repetition variability (thereby improving the reliability and precision of the measurements), the benefit must be weighed against the additional time and resources it takes to acquire more images. Shrout and Fleiss²³ discussed the advantages of using a mean of ratings when the reliability of a single measure was deemed unacceptable. They purported that the number of ratings used to form the mean should be determined by a reliability study. To our knowledge, no such study has been performed investigating the effect on reliability of using a mean of other than 3 RUSI measures.

The primary purpose of this analysis was to investigate the improvements in measurement precision when using the mean value of percent change in muscle thickness of the TrA and LM muscles from multiple repetitions of a standardized task. We hypothesized that measurement precision would increase using the mean of measures, and that there would be a threshold, where averaging additional measurements would not result in large enough gains in precision to warrant the additional effort. The secondary purpose was to assess for a systematic order effect (eg, learning and/or fatigue) when using sequential measurements of percent change in TrA and LM muscle thickness during standardized tasks.

METHODS

THE DATA FROM THIS ANALYSIS COME from a reliability study investigating the intraexaminer and interexaminer reliability in obtaining RUSI thickness measurements of the TrA and LM muscles. The full details of this study have been reported elsewhere.¹⁷ The current analysis concerns only intraexaminer reliability estimates based on the 6 sequential measurements taken by the same examiner during the first imaging session.

Participants

Thirty volunteers, aged 18 to 60 years, with current LBP, with or without lower extremity symptoms, were recruited for this study. Exclusion criteria were prior history of lumbar surgery and presence of medical “red flags” of conditions such as cauda equina syndrome, major or rapidly progressing neurological deficit, fracture, cancer, infection, or systemic disease. Participants provided consent on forms approved by the University of Utah and Intermountain Healthcare Institutional Review Boards, and the rights of all participants were protected.

Examiner

A chiropractor (J.H.) that had not previously used RUSI in clinical practice served as the examiner for this reliability analysis. Prior to testing, he underwent 16 hours of practical training with a co-investigator (D.T.) experienced with the specific RUSI protocol.

Procedures

After signing consent forms, participants completed self-report measures, including demographic and historical information, and questionnaires on pain and disability. An 11-point numeric pain rating scale was used to assess current pain intensity and best and worst pain intensity in the past 24 hours.^{3,13,18} The modified Oswestry questionnaire was used to quantify self-reported disability.⁵ A standard physical examination was then performed to assess for study inclusion



FIGURE 1. An active straight-leg raise (ASLR) maneuver was used to automatically activate the transversus abdominis (TrA) muscle.

and exclusion criteria and to determine the participant's symptomatic side, which was used for all subsequent RUSI images. If pain was evenly distributed, the imaged side was determined randomly.

All images were acquired with a Sonosite Titan ultrasound machine (Sonosite Inc, Bothell, WA) and a 60-mm, 2- to 5-MHz curvilinear array used in B-mode at 5 MHz. Image acquisition for each condition was performed 6 times. The examiner positioned the transducer and optimized the quality of each image, while an assistant captured and saved the image. To control for the influence of respiration, all images were captured at the end of normal exhalation. The order in which the muscles (TrA and LM) were imaged was counterbalanced.

Transversus Abdominis Images of the TrA muscle were acquired in the following order: (1) supine at rest, with lower extremities extended, (2) during an active straight-leg raise (ASLR) maneuver (ipsilateral lower extremity extended), (3) hook lying at rest, and (4) during an abdominal drawing-in maneuver (ADIM) performed in hook lying. Ultrasound images of the TrA muscle were obtained following the techniques outlined by Teyhen et al,²⁹ with the transducer positioned just superior to the iliac crest along the midaxillary line and the middle of the muscle belly centered within the field of view.

The ASLR maneuver (**FIGURE 1**) has been advocated as an assessment for lumbopelvic dysfunction^{19,20} and was used in this study to assess automatic (ie, nonvolitional) changes in muscle thickness of the TrA muscle. The lower extremity con-

[RESEARCH REPORT]



FIGURE 2. An abdominal drawing-in maneuver was used to volitionally activate the transversus abdominis muscle.



FIGURE 3. A contralateral arm lift maneuver was performed to activate the lumbar multifidus muscle.

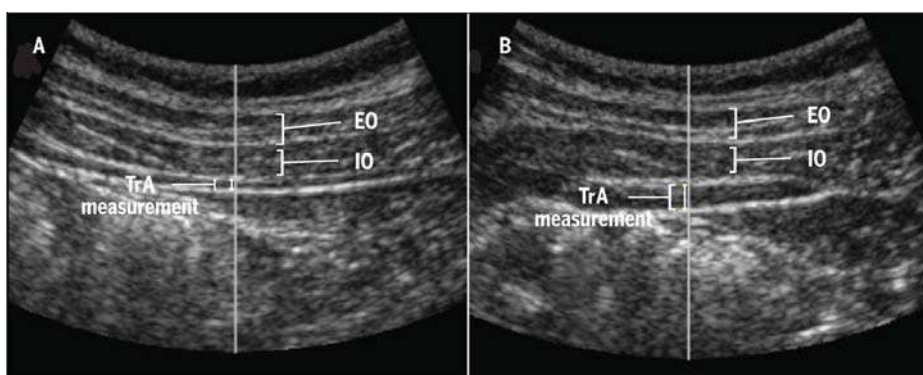


FIGURE 4. Ultrasound image of the transversus abdominis (TrA), internal oblique (IO), and external oblique (EO) muscles during rest (A) and during an abdominal drawing-in maneuver (B). Thickness measurements were made between the superficial and deep borders of the TrA muscle.

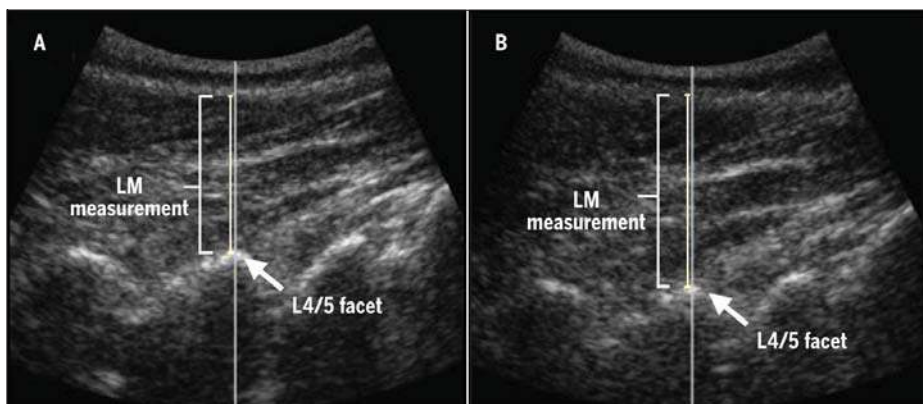


FIGURE 5. Ultrasound image of the lumbar multifidus (LM) muscle during rest (A) and during a contralateral arm raise (B). Thickness measurements were made between the posteriormost portion of the L4-5 facet joint and the plane between the superficial muscle and subcutaneous tissue.

breath in and out, hold the breath out, and then draw-in your lower abdomen without moving your spine.”^{24,29} Alternate cues of “cut off the flow of urine” or “close your rear passage” were sometimes given in an attempt to maximize a preferential TrA contraction. The cue resulting in the largest preferential change in muscle thickness of the TrA during an ADIM was used, as the subject practiced the ADIM until a ceiling effect occurred in performance (approximately 5 times).

Lumbar Multifidus Images of the LM muscle were taken (1) during prone rest and (2) during a contralateral arm raise (CAL) maneuver (FIGURE 3), following techniques outlined by Kiesel et al.¹⁵ The transducer was placed longitudinally (parasagittal view), then angled slightly medially until the facet joints were visualized. After identifying the superior border of the sacrum, the transducer was moved superiorly until the L4-5 facet was centered on the screen. The CAL maneuver was performed prone with the elbows flexed 90°, shoulders abducted 120°, and holding a hand weight (FIGURE 3). The weight used depended on the participant’s body weight and has been shown to elicit approximately 30% of the maximal voluntary isometric contraction of the LM muscle.¹⁵ Participants weighing less than 68 kg used 0.68 kg, those between 68 and 90 kg used 0.91 kg, and those greater than 91 kg used 1.36 kg. Participants were instructed to “lift your arm approximately 2 inches (5 centimeters) off the table” and were given a single practice CAL trial before image acquisition.

Measurements All images were measured offline using National Institutes of Health (Bethesda, MD) Image J, Version 1.38t. TrA thickness measurements were made between the superficial and deep borders of the muscle, as visualized by the hyperechoic fascial lines (FIGURE 4). LM thickness measurements were made between the posteriormost portion of the L4-5 zygapophyseal joint and the plane between the muscle and subcutaneous tissue (FIGURE 5). By using Image J’s automatic measurement function (Ctrl+M) and concealing the measurement output

tralateral to the imaged side was lifted, as it seemed to elicit the same or larger TrA response as the ipsilateral lower extremity during pilot work and was consistent with raising the contralateral arm when testing the LM. Participants were instructed to “raise your leg off of the table approximately 8 inches (20 centimeters), with-

out bending your knee” and were given a single practice of the ASLR maneuver before image acquisition. The ADIM (FIGURE 2) is a motor control exercise for the TrA muscle and was utilized in this study to assess changes in muscle thickness associated with a volitional activation. Participants were instructed to “take a relaxed

TABLE 1

DEMOGRAPHIC AND BASELINE CHARACTERISTICS OF PARTICIPANTS*

Age (y)	42.4 (11.4)
Sex	43% women
Body mass index (kg/m ²)	26.6 (4.8)
Oswestry disability score (%)	20.4 (14.1)
Numeric pain rating scale [†]	2.9 (2.0)
Duration of symptoms (d)	75 (17, 847) [‡]
Prior history of low back pain	80%

* Values are mean (SD) unless otherwise indicated (n = 30).

[†] Reports the average of the worst, best, and current scores for pain over the previous 24 hours, based on a 0-to-10 scale with 0 being no pain and 10 being pain as bad as it can be.

[‡] Median (interquartile range).

TABLE 2

INTRAEXAMINER RELIABILITY

Number of Measurements	ICC _{3,k} (95% CI)	SEM (95% CI)
TrA during an ASLR		
Single	0.83 (0.67, 0.91)	15.1% (11.8%, 19.9%)
Mean of 2	0.90 (0.80, 0.95)	11.2% (8.7%, 14.7%)
Mean of 3	0.95 (0.91, 0.97)	8.1% (6.4%, 10.7%)
Mean of 4	0.93 (0.88, 0.96)	9.4% (7.4%, 12.4%)
Mean of 5	0.95 (0.91, 0.97)	8.4% (6.6%, 11.1%)
Mean of 6	0.95 (0.92, 0.98)	7.8% (6.1%, 10.3%)
TrA during an ADIM		
Single	0.79 (0.61, 0.89)	19.2% (15.0%, 25.4%)
Mean of 2	0.88 (0.75, 0.94)	14.3% (11.2%, 18.9%)
Mean of 3	0.94 (0.89, 0.97)	10.3% (8.0%, 13.6%)
Mean of 4	0.93 (0.88, 0.96)	11.0% (8.6%, 14.5%)
Mean of 5	0.95 (0.92, 0.97)	9.2% (7.2%, 12.1%)
Mean of 6	0.96 (0.94, 0.98)	8.2% (6.4%, 10.8%)
LM during a CAL		
Single	0.88 (0.76, 0.94)	3.0% (2.3%, 4.0%)
Mean of 2	0.94 (0.87, 0.97)	2.2% (1.7%, 2.9%)
Mean of 3	0.96 (0.92, 0.98)	1.8% (1.4%, 2.4%)
Mean of 4	0.92 (0.86, 0.96)	2.4% (1.9%, 3.2%)
Mean of 5	0.92 (0.86, 0.96)	2.5% (1.9%, 3.3%)
Mean of 6	0.93 (0.88, 0.96)	2.3% (1.8%, 3.1%)

Abbreviations: ADIM, abdominal drawing-in maneuver; ASLR, active straight-leg raise; CAL, contralateral arm lift; CI, confidence interval; ICC, intraclass correlation coefficient; LM, lumbar multifidus; SEM, standard error of measurement; TrA, transverse abdominis.

estimated using 2-way mixed-model, consistency-type intraclass correlation coefficients (ICCs). The reliability when using a single measurement was estimated using the first 2 measurement variables and the “single measures” output from SPSS (model 3,1).²³ The reliability when using a mean of 2 measurements was estimated using the first 2 measurement variables and the “average measures” output from SPSS (model 3,2).²³ The reliability, when using a mean of 3, 4, 5, and 6 measurements, was similarly estimated using the first 3, 4, 5, and 6 measurement variables and the “average measures” output from SPSS (model 3,k).²³

Using the resulting ICC values and the equation $SD_{pooled} \times \sqrt{(1 - ICC)}$, SEM was calculated to assess measurement precision.^{21,26} The 95% CIs were estimated for SEM using sum-of-squares error divided by the chi-square value for probability level α (.025 for the lower limit and .975 for the upper limit).²⁵

We visually assessed for the presence of a systematic order effect (eg, learning and/or fatigue) by graphing the mean value of the first, second, third, fourth, fifth, and sixth measurements of percent change in TrA and LM muscle thickness. Repeated-measures analysis of variance (ANOVA) tests were then performed to assess for stability of sequential measurements using $\alpha = .05$ for each comparison.

RESULTS

DEMOGRAPHIC AND BASELINE CHARACTERISTICS of the patient sample are provided in TABLE 1. Images from 1 participant for the LM muscle were excluded because the examiner was unable to identify muscle boundaries, leaving 30 subjects with full TrA images and 29 with full set of LM images. Reliability coefficients and SEM (with 95% CIs), when using a single measurement and when using a mean of 2, 3, 4, 5, and 6 measurements, are presented in TABLE 2. SEM decreased between single measurements,

on the computer screen, the examiner was blinded during measurement to the current and previous thickness values.

Data Analysis

Statistical analysis was performed using the Statistical Package for the Social Sciences, Version 16.0 software (SPSS,

Chicago, IL). The dependent measure of percent change in TrA and LM muscle thickness from rest was calculated as $(\text{thickness}_{\text{contracted}} - \text{thickness}_{\text{rest}}) / \text{thickness}_{\text{rest}}$. Each sequential measurement was arranged in SPSS as a separate variable, and intraexaminer reliability with 95% confidence intervals (CIs) was

a mean of 2 measurements, and a mean of 3 measurements for both the TrA and LM muscles during all standardized tasks (FIGURE 6). Averaging more than 3 measurements resulted in little to no further decrease in SEM.

The mean (SD) change in TrA muscle thickness across all 6 trials during the ASLR was 17.1% (36.0%) and 82.5% (41.9%) during the ADIM. The mean (SD) change in LM muscle thickness across all 6 trials of the CAL was 10.5% (8.7%). No significant differences were found between mean changes in muscle thickness of each sequential measurement for any muscle or standardized task ($P < .05$). Furthermore, no obvious upward (learning) or downward (fatigue) “trends” were apparent from visual inspection of mean percent change in muscle thickness data during any standardized task (FIGURE 7).

DISCUSSION

THE RESULTS FROM THIS ANALYSIS suggest that, when using percent change in muscle thickness as an indirect assessment of muscle activation, using the mean of multiple measures substantially improves intraexaminer measurement precision. Our results suggest that measurement precision is optimized by using an average of 3 consecutive measurements of the TrA and LM during standardized tasks. Regardless of muscle (TrA, LM) or standardized task (ADIM, ASLR, CAL), SEM decreased approximately 25% when using a mean of 2 measures, and by almost 50% when using the mean of 3 measures, when compared to a single measure. Averaging more than 3 measures did not result in consistent additional improvements in measurement precision.

When deciding how many measurements to average, practitioners need to weigh the benefits of increased precision with the costs of spending more time and resources to take each additional measurement. Unfortunately no studies to date have established minimal clinically important differences (MCIDs) of percent change in TrA or LM muscle thick-

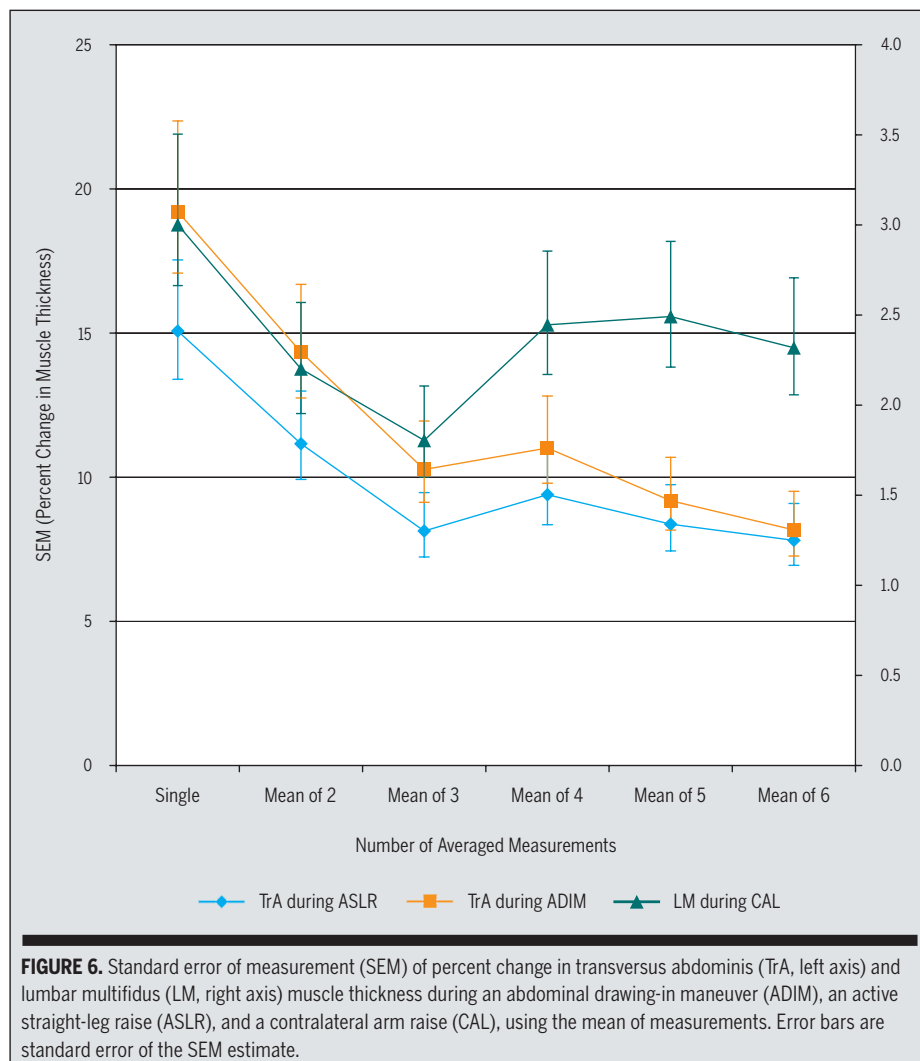


FIGURE 6. Standard error of measurement (SEM) of percent change in transversus abdominis (TrA, left axis) and lumbar multifidus (LM, right axis) muscle thickness during an abdominal drawing-in maneuver (ADIM), an active straight-leg raise (ASLR), and a contralateral arm raise (CAL), using the mean of measurements. Error bars are standard error of the SEM estimate.

ness. However, the improvements found in SEM when averaging 2 measurements as compared to a single measurement, and 3 measurements as compared to 2 measurements, are both at least moderate in relation to mean percent thickness change values. Because the cost of taking each additional measurement is minimal (perhaps 1-2 minutes to capture and measure each additional image), using the mean of 3 measurements seems to provide the optimal cost-benefit ratio.

Although improvements in measurement precision cannot be compared to MCIDs, they can be compared to differences reported between symptomatic and asymptomatic individuals. The mean percent change in muscle thickness in this symptomatic sample was 17.1% for the TrA during the ASLR, 82.5% for the

TrA during the ADIM, and 10.5% for the LM during the CAL. In asymptomatic individuals, Kiesel et al¹⁶ reported a mean TrA percent change in muscle thickness of 99.1% during an ADIM and a mean LM percent change in L4-5 muscle thickness of 24.4% during a CAL. Comparison of the symptomatic and asymptomatic values between the current study and that of Kiesel et al¹⁶ suggests that patients with LBP (compared to persons who are asymptomatic) demonstrate a diminished change in muscle thickness of the LM, at a level that is greater than measurement error (SEM), even when using only a single measure. In terms of percent change in the TrA muscle during the ADIM, the difference between persons who are symptomatic and those who are asymptomatic is greater than measurement er-

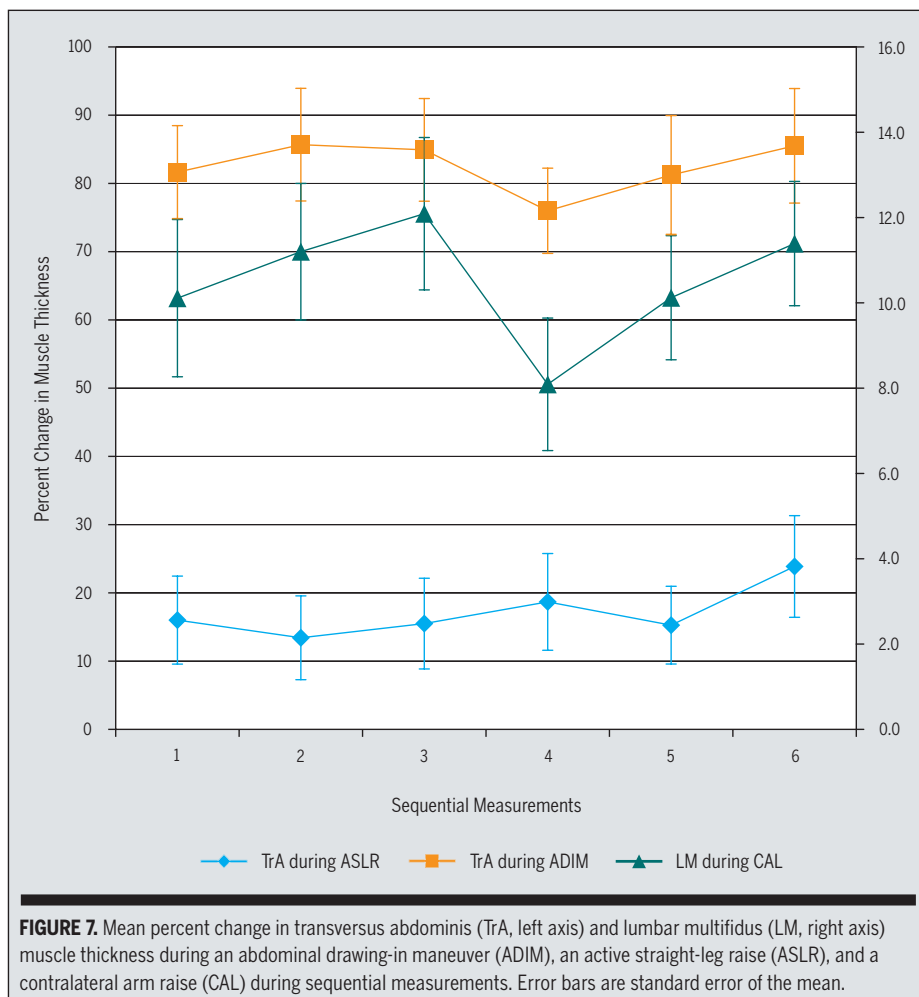


FIGURE 7. Mean percent change in transversus abdominis (TrA, left axis) and lumbar multifidus (LM, right axis) muscle thickness during an abdominal drawing-in maneuver (ADIM), an active straight-leg raise (ASLR), and a contralateral arm raise (CAL) during sequential measurements. Error bars are standard error of the mean.

ror (SEM) only when using the mean of at least 2 measurements.

The current study also has implications for the responsiveness of RUSI measures during rehabilitation. MDC is proportional to, although larger than, SEM (95% MDC = $1.96 \times \text{SEM} \times \sqrt{2}$). To exceed the MDC found in this study, an average patient with LBP would have to increase the percent change in LM muscle thickness from 10.5% to 18.8% if a single measure were used, and from 10.5% to 15.5% if a mean of 3 measures were used. Similarly, a patient would have to increase the percent change in TrA muscle thickness during an ADIM from 82.5% to 135.0% if a single measure were used, and from 82.5% to 111.0% if a mean of 3 measures were used. So, while using the mean of 3 measures may help monitor percent change in LM muscle thickness in patients with LBP during rehabilitation,

those patients would have to increase their percent change in TrA muscle thickness to exceed that seen in persons who are asymptomatic before we could be confident that a real change occurred.

Unfortunately, no other study to our knowledge has reported percent change in TrA muscle thickness during an ASLR. So, although we don't know the mean percent change in muscle thickness in asymptomatic individuals, the SEM found in this study is large in relation to the 17.1% change in muscle thickness found in this symptomatic sample. Specifically, measurement error would be 15.1% if a single measure were used and 8.1% were a mean of 3 measures used. The large amount of error within TrA measurements during both the ADIM and ASLR is likely related to the high degree of intersubject variability (large SDs) during both standardized tasks. This suggests

the need for identifying more reliable contraction strategies of the TrA, as well as methods to further reduce error during such measurements.

In the primary reliability analysis of this study¹⁷ we focused on the reliability of thickness measures as opposed to percent change in muscle thickness measures. We decided to report our primary results using the mean of 2 measures, as there seemed to be very little additional benefit to averaging in a third measure. Single thickness measures may be important if used to assess within person changes over time, when one expects muscle atrophy or hypertrophy. Although some studies have found LM muscle atrophy in patients with LBP,^{7,33} most of the relevant literature focuses on functional deficits of the TrA and LM muscles.^{1,2,4,10-12,30} Because thickness measures likely relate more to a person's size than their muscle function, percent change in muscle thickness is arguably a more clinically relevant measure than that of thickness. However, as expected, measurement of percent change is less reliable than single thickness measures, as it incorporates measurement error from both resting and contracted states. This is evident from our analysis and is consistent with the recommendation of using the mean of 3 measurements as opposed to 2 measurements when using percent change in muscle thickness as the dependent variable.

When using the mean of multiple measures to best estimate an individual's true measure, it is important not to have a systematic order effect. Because muscle performance measures are naturally prone to be influenced by fatigue and/or learning effects, the second part of this analysis was aimed at identifying any systematic bias related to measurement order. Regardless of muscle (TrA, LM) or standardized task (ADIM, ASLR, CAL), we found that 6 consecutive measures, when taken within approximately 5 to 10 minutes, are relatively stable. The lack of an identifiable upward or downward trend in sequential means (FIGURE 7) suggests that neither fatigue nor learning systematically affected percent change in

TrA or LM muscle thickness during any standardized task.

In addition to the limitations discussed in our primary reliability report,¹⁷ this analysis is limited by the inclusion of only intraexaminer data obtained during the same day. While we believe that this analysis yields important information for RUSI practitioners and researchers, we cannot generalize the results to measurement error when images are taken during different days or by different examiners. Because making between-day and/or between-examiner comparisons add additional error, it is possible that such comparisons would benefit from more than 3 measurements. Moreover, quantitative interpretations of measurement error would ideally be made by comparing changes in SEM to established MCIDs. Because MCIDs have not been established for any RUSI measurements, we cannot comment on the importance of the reported improvements in measurement precision. Lastly, error during measurement can be due to multiple sources, including the instrument, the examiner (eg, transducer motion), and intra-subject variability. Identifying the amount of measurement error that comes from each of these different sources requires a much larger sample and a different analysis (generalizability theory) and cannot be inferred from this study.

Future studies should continue to investigate measurement error associated with RUSI, especially using the more clinically relevant measure of percent change in muscle thickness during muscle activation. Because absolute measurement error (SEM) is relatively large in comparison to mean TrA measures, studies should focus on identifying strategies to improve the reliability of both acquiring and measuring TrA images.

CONCLUSION

WHEN USING RUSI TO DETERMINE percent change in TrA and LM muscle thickness, intraexaminer measurement precision is substantially improved by using the mean of multiple

measures. Using an average of 3 measurements of the TrA and LM muscles appears to optimize measurement precision and is recommended. ●

KEY POINTS

FINDINGS: When using RUSI to determine percent change in TrA and LM muscle thickness, intraexaminer measurement precision was optimized by using an average of 3 consecutive measurements. Little precision was gained by averaging more than 3 measurements.

IMPLICATION: Using an average of 3 measurements when assessing percent thickness change of the TrA and LM muscles substantially improves measurement precision compared to fewer trials and is recommended for both research and clinical use.

CAUTION: This analysis is limited by the inclusion of only intraexaminer data obtained during the same day. Comparisons using measurements from different days and from different examiners include additional error and, therefore, may benefit from more than 3 measurements. Moreover, because MCIDs have not been established for RUSI measurements, this analysis cannot evaluate the “importance” of these improvements in measurement precision or allow us to identify specific sources of measurement error.

REFERENCES

- Critchley D. Instructing pelvic floor contraction facilitates transversus abdominis thickness increase during low-abdominal hollowing. *Physiother Res Int*. 2002;7:65-75.
- Critchley DJ, Coultts FJ. Abdominal muscle function in chronic low back pain patients: measurement with real-time ultrasound scanning. *Physiotherapy*. 2002;88:322-332.
- Farrar JT, Berlin JA, Strom BL. Clinically important changes in acute pain outcome measures: a validation study. *J Pain Symptom Manage*. 2003;25:406-411.
- Ferreira PH, Ferreira ML, Hodges PW. Changes in recruitment of the abdominal muscles in people with low back pain: ultrasound measurement of muscle activity. *Spine*. 2004;29:2560-2566.
- Fritz JM, Irrgang JJ. A comparison of a modified Oswestry Low Back Pain Disability Questionnaire and the Quebec Back Pain Disability Scale. *Phys Ther*. 2001;81:776-788.
- Hides JA, Mirkovic T, Belavy DL, Stanton WR, Richardson CA. Ultrasound imaging assessment of abdominal muscle function during drawing-in of the abdominal wall: an intrarater reliability study. *J Orthop Sports Phys Ther*. 2007;37:480-486. <http://dx.doi.org/10.2519/jospt.2007.2416>
- Hides JA, Stokes MJ, Saide M, Jull GA, Cooper DH. Evidence of lumbar multifidus muscle wasting ipsilateral to symptoms in patients with acute/subacute low back pain. *Spine*. 1994;19:165-172.
- Hodges PW. Changes in motor planning of feed-forward postural responses of the trunk muscles in low back pain. *Exp Brain Res*. 2001;141:261-266. <http://dx.doi.org/10.1007/s002210100873>
- Hodges PW. Is there a role for transversus abdominis in lumbo-pelvic stability? *Man Ther*. 1999;4:74-86. <http://dx.doi.org/10.1054/math.1999.0169>
- Hodges PW, Richardson CA. Altered trunk muscle recruitment in people with low back pain with upper limb movement at different speeds. *Arch Phys Med Rehabil*. 1999;80:1005-1012.
- Hodges PW, Richardson CA. Delayed postural contraction of transversus abdominis in low back pain associated with movement of the lower limb. *J Spinal Disord*. 1998;11:46-56.
- Hungerford B, Gilleard W, Hodges P. Evidence of altered lumbopelvic muscle recruitment in the presence of sacroiliac joint pain. *Spine*. 2003;28:1593-1600.
- Jensen MP, Turner JA, Romano JM, Fisher LD. Comparative reliability and validity of chronic pain intensity measures. *Pain*. 1999;83:157-162.
- Kidd AW, Magee S, Richardson CA. Reliability of real-time ultrasound for the assessment of transversus abdominis function. *J Gravit Physiol*. 2002;9:P131-132.
- 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. <http://dx.doi.org/10.1016/j.math.2006.06.011>
- Kiesel KB, Underwood FB, Mattacola CG, Nitz AJ, Malone TR. A comparison of select trunk muscle thickness change between subjects with low back pain classified in the treatment-based classification system and asymptomatic controls. *J Orthop Sports Phys Ther*. 2007;37:596-607. <http://dx.doi.org/10.2519/jospt.2007.2574>
- 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. <http://dx.doi.org/10.1016/j.apmr.2008.06.022>
- Li L, Liu X, Herr K. Postoperative pain intensity assessment: a comparison of four scales in Chinese adults. *Pain Med*. 2007;8:223-234. <http://dx.doi.org/10.1111/j.1526-4637.2007.00296.x>
- Mens JM, Vleeming A, Snijders CJ, Koes BW, Stam HJ. Validity of the active straight leg raise test for measuring disease severity in patients

with posterior pelvic pain after pregnancy. *Spine*. 2002;27:196-200.

20. Mens JM, Vleeming A, Snijders CJ, Stam HJ, Ginai AZ. The active straight leg raising test and mobility of the pelvic joints. *Eur Spine J*. 1999;8:468-473.
21. Portney LG, Watkins MP. *Foundations of Clinical Research: Applications to Practice*. 3rd ed. Upper Saddle River, NJ: Pearson Prentice Hall; 2008.
22. 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. <http://dx.doi.org/10.2519/jospt.2006.2049>
23. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86:420-428.
24. Springer BA, Mielcarek BJ, Nesfield TK, Teyhen DS. Relationships among lateral abdominal muscles, gender, body mass index, and hand dominance. *J Orthop Sports Phys Ther*. 2006;36:289-297. <http://dx.doi.org/10.2519/jospt.2006.2217>

jospt.2006.2217

25. Stratford PW, Goldsmith CH. Use of the standard error as a reliability index of interest: an applied example using elbow flexor strength data. *Phys Ther*. 1997;77:745-750.
26. Streiner DL, Norman GR. *Health Measurement Scales: A Practical Guide to Their Development and Use*. New York, NY: Oxford University Press; 2003.
27. Teyhen D. Rehabilitative Ultrasound Imaging Symposium San Antonio, TX, May 8-10, 2006. *J Orthop Sports Phys Ther*. 2006;36:A1-3. <http://dx.doi.org/10.2519/jospt.2006.0301>
28. Teyhen DS, Gill NW, Whittaker JL, Henry SM, Hides JA, Hodges P. Rehabilitative ultrasound imaging of the abdominal muscles. *J Orthop Sports Phys Ther*. 2007;37:450-466. <http://dx.doi.org/10.2519/jospt.2007.2558>
29. Teyhen DS, Miltenberger CE, Deiters HM, et al. The use of ultrasound imaging of the abdominal drawing-in maneuver in subjects with low back pain. *J Orthop Sports Phys Ther*. 2005;35:346-355. <http://dx.doi.org/10.2519/jospt.2005.1780>
30. 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. <http://dx.doi.org/10.2519/jospt.2006.2304>

31. 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. <http://dx.doi.org/10.2519/jospt.2007.2418>
32. Yoshihara K, Shirai Y, Nakayama Y, Uesaka S. Histochemical changes in the multifidus muscle in patients with lumbar intervertebral disc herniation. *Spine*. 2001;26:622-626.
33. Zhao WP, Kawaguchi Y, Matsui H, Kanamori M, Kimura T. Histochemistry and morphology of the multifidus muscle in lumbar disc herniation: comparative study between diseased and normal sides. *Spine*. 2000;25:2191-2199.



MORE INFORMATION
WWW.JOSPT.ORG

EARN CEUs With JOSPT's Read for Credit Program

JOSPT's **Read for Credit (RFC)** program invites *Journal* readers to study and analyze selected *JOSPT* articles and successfully complete online quizzes about them for continuing education credit. To participate in the program:

1. Go to www.jospt.org and click on **"Read for Credit"** in the left-hand navigation column that runs throughout the site or on the link in the **"Read for Credit"** box in the right-hand column of the home page.
2. Choose an article to study and when ready, click **"Take Exam"** for that article.
3. Login and pay for the quiz by credit card.
4. Take the quiz.
5. Evaluate the RFC experience and receive a personalized certificate of continuing education credits.

The RFC program offers you 2 opportunities to pass the quiz. You may review all of your answers—including the questions you missed. You receive **0.2 CEUs**, or 2 contact hours, for each quiz passed. The *Journal* website maintains a history of the quizzes you have taken and the credits and certificates you have been awarded in the **"My CEUs"** section of your **"My JOSPT"** account.