Ultrasound Imaging Assessment of Abdominal Muscle Function During Drawing-in of the Abdominal Wall: An Intrarater Reliability Study

Assessment of function of the deep abdominal musculature in the laboratory has largely used methodologies such as functional movement tasks and fine-wire electromyography (EMG). While fine-wire EMG is primarily restricted to laboratory research, rehabilitative ultrasound imaging (RUSI) offers a noninvasive and simpler method of assessing deep lumbopelvic muscle function. Recent work has shown that measures of transversus abdominis (TrA) muscle contraction, internal oblique (IO) muscle contraction, and length of the TrA muscle obtained using RUSI correlated well with measures obtained by fine-wire EMG for isometric contractions of these muscles. For increases in thickness of the TrA and IO muscles and changes in length of the TrA muscle (measured as lateral displacement of the fascial insertion of the TrA muscle) there were consistent, clear changes in sonographic parameters, with incremental changes for contractions of less than 20% of maximal voluntary contraction (MVC).

Clinical muscle testing of the TrA muscle has been based on its anatomical structure and horizontal fiber arrangement. Two muscle tests include observation of the abdominal wall during either a voluntary drawing-in of the abdominal wall or by using automatic responses of the muscle to expiration. A relationship between the clinical muscle test and laboratory EMG measures of TrA muscle function has been demonstrated. In addition, the action of the bilateral muscle bellies of the TrA muscle has been viewed during the drawing-in maneuver using RUSI and magnetic resonance imaging (MRI). In a recent study conducted on elite asymptomatic cricketers, the muscle bellies of the TrA muscle were seen to thicken as well as shorten in length during...
ing this maneuver, to give the appearance of a deep muscle “corset,” as indicated by anatomical studies. Measurements of the lateral displacement of the anterior facial insertion of the TrA muscle and of the thickness of the TrA and IO muscles at rest and on contraction were conducted using both RUSI and MRI. Measurements obtained using RUSI correlated well with measurements obtained using MRI [intraclass correlation coefficients (ICCs) ranging from 0.78 to 0.95]. Differences in patterns of activation of the abdominal muscles during the drawing-in maneuver have been documented in subjects with low back pain (LBP) using MRI. Subjects with LBP were shown to be less able than asymptomatic subjects to draw-in the abdominal wall and also displayed significantly less lateral slide of the anterior abdominal fascia.

Randomized clinical trials that have focused on re-educating TrA muscle function through its drawing-in action have been successful in decreasing lumbar pelvic pain and disability. Goldby et al. successfully used RUSI to provide feedback of abdominal muscle activation to subjects with chronic LBP. In further support of its use, a recent randomized controlled trial showed that using RUSI to provide feedback of TrA muscle activation was superior to clinical instruction alone in healthy subjects.

While RUSI is being used increasingly to examine the abdominal muscles in clinical practice, there are only a few studies that have examined reliability of measurement of these muscles. To date, researchers have found RUSI to be reliable for measuring thickness of the abdominal muscles at rest and changes in thickness of the abdominal muscles when they contract. Although intrarater reliability of 2 measurers (experienced therapist and a novice) has been shown, studies have not, as yet, reported novice therapists’ intrarater reliability across a range of measurement conditions. The purpose of this study, which used RUSI to assess the TrA and IO muscles on both sides at rest and on contraction, was to examine measurements made by a physical therapist who had little experience in RUSI but underwent a short, supervised training program.

**MATERIALS AND METHODS**

**Subjects**

**Nineteen normal healthy subjects (11 female, 8 male)** participated in the study. The mean (±SD) age, height, and body mass of the subjects was 20.3 ± 5.0 years, 172.0 ± 9.8 cm, and 64.5 ± 11.4 kg, respectively. Exclusion criteria were a history of LBP, previous lumbar surgery, known neuromuscular or joint disease, significant spinal abnormality (eg, scoliosis), participation in competitive sports more than 3 times a week, pregnancy, and familiarity with the testing procedure. An asymptomatic population was selected, as it was assumed that their performance of the abdominal drawing-in maneuver would be relatively consistent on the 2 testing days. The study was approved by the Medical Research Ethics Committee at The University of Queensland, Australia. Informed consent was obtained from each subject and the rights of human subjects were protected.

**Instrumentation and Training**

The anterolateral abdominal muscles were imaged in brightness mode (b-mode) on both sides, using a Synergy ultrasound apparatus (GE-Diasonics, San Jose, CA). A physical therapist underwent training and performed all of the subsequent measurements. Training in RUSI of the anterolateral abdominal muscles included 2 hours of 1-on-1 practice supervised by the senior researcher in this project. In addition, the physical therapist had received 3 hours of lectures and 3 hours of practical instruction on RUSI as part of the physical therapy course curriculum.

**Subject Familiarization With the Testing Procedure** Subjects were positioned in a supine hook-lying position, with the hips in 45° of flexion (Figure 1). A pressure biofeedback unit (Chattanooga, Hixson, TN), a device that detects changes in pressure, was placed under the subject’s lumbar spine and inflated to 40 mmHg. Subjects were then provided with standard instructions on how to activate their TrA muscle using the drawing-in maneuver: “Take a relaxed breath in and out, hold the breath out, and then draw in your lower abdomen without moving your spine.” During this maneuver, the pressure biofeedback unit was monitored for any pressure increase indicative of spinal movement. If spinal movement was detected (suggestive of activation of the rectus abdominis or oblique abdominal muscles to induce posterior pelvic tilt), the subjects were asked to repeat the activation without moving the spine, to try to activate the TrA muscle.

**Testing Protocol** For RUSI assessment, subjects were placed in the same position as the familiarization procedure. During testing, subjects performed 6 drawing-in maneuvers, separated by a 2-minute rest between each. During these maneuvers, RUSI of the anterolateral abdominal wall was performed (3 times each on the left and right sides) in a randomized order. A transverse image of the anterolateral abdominal wall was obtained along a line midway between the inferior angle of the rib cage and the iliac crest for left and right sides. The ultrasound transducer was aligned perpendicular to the anterolateral abdominal muscles. To

**FIGURE 1.** The subject was positioned in a supine hook-lying position, with the hips in 45° of flexion. The ultrasound transducer was aligned perpendicular to the anterolateral abdominal wall, along a line midway between the inferior angle of the rib cage and the iliac crest. Measurements were made for the left and right sides.
standardize the position of the transducer, the anterior fascial insertion of the TrA muscle was positioned approximately 2 cm from the medial edge of the ultrasound image when the subject was relaxed. The subject was not able to see the ultrasound monitor, so did not receive feedback of performance of the contraction. Ultrasound images were obtained at rest, with the subjects instructed to “hold the breath out” and during contraction, while holding the drawing-in maneuver. Images were stored for offline analysis (Figure 2).

Intrarater reliability of RUSI measurements for a novice assessor was examined (a) across 3 repeated measurements of the same image in a rest and contracted condition (using the right-sided image from the first 10 subjects), (b) across 3 recaptured images taken for each side of the abdomen, at rest and contracted, and (c) across 2 days, 4 to 7 days apart. Image Analysis Still ultrasound images of the muscles were extracted offline. Image visualization and measurements were conducted using a software package (ImageJ, Version 1.36b). On each image, thickness of the TrA and IO muscles was measured at rest and on contraction. The thickness of the external oblique muscle was not assessed in this study, as a previous study showed that thickness changes measured using RUSI did not correlate with activation of this muscle assessed using fine-wire EMG. Measurements on ultrasound images were performed as follows (Figure 2):

- Thickness of the TrA and IO muscles at rest and on contraction (left and right). Thickness of the TrA and IO muscles were measured as the distance between the superior and inferior hyperechoic muscle fascias, at the middle of the image. Measurements were conducted perpendicular to the muscle fascias. Separate images were measured for the rest and contracted conditions.
- Slide of the anterior abdominal fascia (left and right). For measurements of slide of the anterior abdominal fascia, the distance from the medial edge of the TrA muscle to the medial edge of the ultrasound image was measured at rest (as per Hodges et al). This starting position was then superimposed on the contracted image, and the distance from this point to the medial edge of the contracted TrA muscle was measured.

Muscle thicknesses and slide of the TrA fascia were measured in millimeters. All still images were deidentified. Statistical Analysis Intrarater reliability was examined by (1) analysis of variance (ANOVA) to examine systematic change in scores across repeated measurements, using linear mixed-effects models in the R, Version 2.1.1, statistical package (The R Foundation, Wien, Austria); (2) calculation of the ICCs based on components of variance to examine change in the rank order of scores; and (3) calculation of the standard error of measurement (SEM) to examine the precision of measurement (SEM = pooled SD × [1 – ICC]). These analyses were performed for each of the outcome variables: IO muscle thickness (at rest and contracted), TrA muscle thickness (at rest and contracted), and slide of the anterior abdominal fascia. ICCs in the range of 0.0 to 0.5 were considered very low, those ranging from greater than 0.5 to 0.7 were considered low, those greater than 0.7 to 0.9 were considered high, and those greater than 0.9 were considered very high.

Repeated Measures of the Same Image For each outcome variable, analysis of reliability across 3 repeated measurements of the same image was conducted on the first 10 subjects, using the first image taken on the right side. The ANOVA contained this repeated measure, as well as the covariates of age, gender, height, and body mass. For each outcome variable, ICC was calculated. Repeated Images of the Same Object For each outcome variable, analysis across the 3 repeated images was based on an ANOVA design with fixed effects for image, side of abdomen, and 2-way interactions between these factors. Age, gender, height, and body mass were included as covariates in the ANOVA. For each outcome variable, ICC was calculated using the average of the first image taken for both sides on both days (4 images), compared to the second and third set of images taken.

For each outcome variable, analysis of reliability across 3 repeated measurements of the same image was conducted on the first 10 subjects, using the first image taken on the right side. The ANOVA contained this repeated measure, as well as the covariates of age, gender, height, and body mass. For each outcome variable, ICC was calculated. Repeated Images of the Same Object For each outcome variable, analysis across the 3 repeated images was based on an ANOVA design with fixed effects for image, side of abdomen, and 2-way interactions between these factors. Age, gender, height, and body mass were included as covariates in the ANOVA. For each outcome variable, ICC was calculated using the average of the first image taken for both sides on both days (4 images), compared to the second and third set of images taken.

Repeated Measures of the Same Image For each outcome variable, analysis of reliability across 3 repeated measurements of the same image was conducted on the first 10 subjects, using the first image taken on the right side. The ANOVA contained this repeated measure, as well as the covariates of age, gender, height, and body mass. For each outcome variable, ICC was calculated. Repeated Images of the Same Object For each outcome variable, analysis across the 3 repeated images was based on an ANOVA design with fixed effects for image, side of abdomen, and 2-way interactions between these factors. Age, gender, height, and body mass were included as covariates in the ANOVA. For each outcome variable, ICC was calculated using the average of the first image taken for both sides on both days (4 images), compared to the second and third set of images taken.
Repeated Procedure Across Days

The effect across the 2 days was examined by including the fixed effect of day (testing day) in the ANOVAs described above. For each outcome variable, ICC$_{3,6}$ was calculated using the average value of the 6 images taken on day 1, compared to the average value of the 6 images on day 2.

RESULTS

The means and standard deviations of muscle thickness and lateral slide of the anterior abdominal fascia, while performing a drawing-in maneuver, are shown in Table 1.

Repeated Measurements of the Same Image Intrarater reliability was very high for this procedure. The ANOVA results showed no systematic difference in scores for remeasurement of the same image (for all: $F_{20,9}$, 2.38, $P$.12). The single-measure ICC values of the novice assessor for IO and TrA muscle thickness and lateral slide of the anterior abdominal fascia were all greater than 0.97 (Table 2). SEM values less than 0.03 mm help to confirm the accuracy of the rater’s measurements of the same image.

Repeated Images

No systematic difference in scores was found across images in mean IO muscle thickness or TrA muscle thickness measurements at both rest and contracted conditions, or in slide of the anterior abdominal fascia (for all: $F_{5,36}$, <1.10, $P$.30). The ICC values of muscle measurements by the novice assessor ranged from low to high and were not consistent for the rest and contracted conditions (Table 3). The ICC value for the novice’s measurements of slide of the anterior abdominal fascia was very low. The precision of the measurements for recaptured images was lower (indicated by relatively higher SEMs) compared with repeated measurement of the same image. In the case of slide of the anterior abdominal fascia, the lack of precision and wide 95% confidence interval (CI) confirm the poor reliability of the assessor for this procedure.

<table>
<thead>
<tr>
<th>TABLE 1</th>
<th>Thickness Measures of the Internal Oblique and Transversus Abdominis Muscles, and Slide of the Anterior Abdominal Fascia*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measure</td>
<td>Rest</td>
</tr>
<tr>
<td>IO thickness (mm)</td>
<td>7.2 (1.5)</td>
</tr>
<tr>
<td>TrA thickness (mm)</td>
<td>3.4 (0.8)</td>
</tr>
<tr>
<td>TrA slide (mm)</td>
<td>_</td>
</tr>
</tbody>
</table>

Abbreviations: IO, internal oblique muscle; TrA, transversus abdominis muscle.

* Values represent mean (standard deviation) for each dependent measure based on state (rest, contracted), and side (left and right). For each subject, the average of 6 images (3 images obtained on each of 2 days) was used; n = 19.

<table>
<thead>
<tr>
<th>TABLE 2</th>
<th>Intrarater Reliability Across Repeated Measurement of the Same Image*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measures</td>
<td>ICC$_{3,1}$</td>
</tr>
<tr>
<td>IO thickness</td>
<td>0.99 (0.97-1.00)</td>
</tr>
<tr>
<td>TrA thickness</td>
<td>0.98 (0.95-1.00)</td>
</tr>
<tr>
<td>TrA slide</td>
<td>_</td>
</tr>
</tbody>
</table>

Abbreviations: ICC, intraclass correlation (95% confidence interval); IO, internal oblique muscle; SEM, standard error of measurement; TrA, transversus abdominis muscle.

* n = 10.

<table>
<thead>
<tr>
<th>TABLE 3</th>
<th>Intrarater Reliability Across 3 Images*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measures</td>
<td>ICC$_{3,4}$</td>
</tr>
<tr>
<td>IO thickness</td>
<td>0.82 (0.55-0.95)</td>
</tr>
<tr>
<td>TrA thickness</td>
<td>0.62 (0.32-0.85)</td>
</tr>
<tr>
<td>TrA slide</td>
<td>_</td>
</tr>
</tbody>
</table>

Abbreviations: ICC, intraclass correlation coefficient (95% confidence interval); IO, internal oblique muscle; SEM, standard error of measurement; TrA, transversus abdominis muscle.

* The data for each of the 3 images are based on the average of 4 measures (1 image from each side of the abdomen captured on 2 days); n = 19.

<table>
<thead>
<tr>
<th>TABLE 4</th>
<th>Intrarater Reliability Across 2 Days</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measures</td>
<td>ICC$_{3,6}$</td>
</tr>
<tr>
<td>IO thickness</td>
<td>0.69 (0.30-0.92)</td>
</tr>
<tr>
<td>TrA thickness</td>
<td>0.85 (0.42-0.98)</td>
</tr>
<tr>
<td>TrA slide</td>
<td>_</td>
</tr>
</tbody>
</table>

Abbreviations: ICC, intraclass correlation coefficient (95% confidence interval); IO, internal oblique muscle; SEM, standard error of measurement; TrA, transversus abdominis muscle.

* The data for each day are based on the average of 6 measures (3 images taken on each side of the abdomen); n = 19.
Repeated Days  No systematic difference in scores existed across test days in mean IO muscle thickness measurements or TrA muscle thickness at both rest and contracted conditions (for all: $F_{1,36}<2.22$, $P>0.15$). Slide of the anterior abdominal fascia did not differ across test days ($F_{1,36}=1.45$, $P=0.24$). The ICC values of the novice assessor’s IO muscle thickness measurement across day were low, whereas ICC values for TrA muscle measurement were high (Table 4). The ICC value for the novice’s measurement of slide of the anterior abdominal fascia was very low (Table 4). The SEMs showed the greatest variability across day, compared to repeated measurement of same image and recaptured image within the same day.

**DISCUSSION**

Results from this study relating to the morphology of the muscles of the anterolateral abdominal wall can be compared with values from other studies. Rankin et al\(^{24}\) reported values (at rest) for muscle thickness for 55 men and 68 women of various ages who were moderately active. The thicknesses of both the TrA and IO muscles at rest were thicker than values reported in this current study. Hides et al\(^{2}\) reported muscle thicknesses for young asymptomatic male elite cricketers and, not unexpectedly, elite male athletes had thicker muscles at rest than the subjects of the current investigation and the moderately active subjects of Rankin et al.\(^{24}\) One possible explanation for this is muscle hypertrophy in the athletes due to strenuous training and competition. Also, there were females included in the current investigation and only male subjects in the study of Hides et al.\(^{2}\) A gender effect has been previously reported, and male subjects were shown to have significantly thicker muscles than female subjects.\(^{2,24,28}\)

The current investigation examined many aspects of reliability of measurement. Two muscles were measured on both sides of the abdomen in both relaxed and contracted conditions. Furthermore, repeated measurements were conducted from the same stored images, as well as across 3 images and across 2 days. It would be expected that measuring muscles repeatedly from stored images would be associated with the least potential for measurement error. This is because measuring the thickness of the abdominal muscles, which are well defined on the ultrasound image, is a relatively straightforward task. However, accurately reimaging the subject to obtain comparable images may require a higher level of skill. Factors such as relocation of the original imaging site, reproduction of the same transducer pressure and orientation, as well as maintenance of these factors during muscle contraction could adversely affect reliability.

In terms of repeated measurement of the same stored image, the results of a study by Teyhen et al\(^{30}\) and the current study support this premise, with both studies reporting high ICC values and very low SEMs. In the study by Teyhen et al,\(^{30}\) a novice assessor achieved high intrarater reliability for measurement of 2 ultrasound images of the TrA muscle (and a combined measure of the anterolateral abdominal muscles).

In terms of measurement of recaptured images and repetition across days, the novice physical therapist showed variable reliability for attainment of images and subsequent measurement of the thicknesses of the TrA and IO muscles both at rest and on contraction. It may have been expected that ICC values would be higher for measurements of the muscles at rest, as performance of muscle contraction may be inherently variable in nature. Results from the current study showed this to be the case for the IO muscle at rest; but in the case of the TrA muscle, higher ICC values were found for the contracted condition. Measurement precision based on the size of the SEMs showed a similar pattern of variation across these conditions. Other studies have also provided evidence of similar inconsistencies. Misuri et al\(^{11}\) reported coefficients of variation ranging from 0% to 15.7% for intrarater reliability of measures of the TrA and IO muscles at rest and on contraction. Measurements were not always more reliable at rest, with higher reliability estimates reported for the contracted condition in one third of the sample.\(^{21}\)

The amount of training undertaken may also be an important factor when considering reliability of the rater. In their investigation, Ferreira et al\(^{3}\) reported an intrarater ICC of 0.85 for an assessor trained for 3 months in measurement of TrA muscle thickness. This was compared with an assessor with no previous training whose measurements were not reliable (ICC = 0.26).

Across-day reliability may be of interest to physical therapy practitioners who perform repeated assessments of abdominal muscle function over time and who might compare measurements from different treatment sessions. Bunce et al\(^{3}\) measured the thickness of the TrA muscle for 3 different conditions using m-mode and reported across-day intrarater reliability (ICC\(_{1,1}=0.88-0.94\); SEM, 0.35-0.66 mm). When these results are compared with those from the current investigation, the results for our novice assessor were again inconsistent, with better reliability across days for measurement of the TrA muscle both at rest and on contraction than for the IO muscle. In contrast, in a study conducted by Rankin et al,\(^{24}\) ICC values for across-day reliability of measurements of the thickness of the TrA, IO, and external oblique muscles (at rest), conducted by an expert, were consistently high and ranged from 0.96 to 0.99.

The novice assessor in the current investigation was also inconsistent in relation to measurement of slide of the anterior abdominal fascia. The novice was reliable when measuring from stored images but was not able to reliably assess the slide of the anterior abdominal fascia across 3 measurements and across days. As measurement of slide of the anterior abdominal fascia may reflect relative independence in activity of the TrA muscle during the drawing-in maneuver, this measurement may represent a clinically
relevant addition to the more commonly performed measurement of muscle thickness. Furthermore, this measurement reflects that the contraction of the muscle is concentric in nature (ie, the length of the muscle decreases). It is, however, a more difficult measure to perform. This is primarily because it may require more precision and experience to perform the imaging technique well. If the assessor moves the transducer medially or laterally during the draw-in maneuver, there is potential for measurement error. Medial lateral movement of the transducer will not affect thickness measurements to the same degree, as the thickness of the anterolateral abdominal muscles is relatively constant and the direction of the thickness measurements is in an anteroposterior direction. These measurements will not vary greatly if the transducer is moved slightly laterally or medially. However, movement of the transducer in either a medial or lateral direction will have a large impact on measurements conducted in the medial lateral direction. This movement of the transducer can be quite difficult to control, as the contours of the abdominal wall may change when the muscles contract during the abdominal drawing-in maneuver. Placing the transducer in a dense foam cube, as in the procedure used by Ferreira et al,4 may improve the reliability of the assessor at performing this measurement. Future studies are required to determine how much training of assessors is required to reliably perform measurement of slide of the anterior abdominal fascia during the drawing-in maneuver.

The limitations of this study include a relatively small sample, use of only 1 novice rater, and only healthy subjects being investigated. It may be more difficult to reliably measure subjects with symptoms as their performance of the abdominal drawing-in maneuver can be less consistent on 2 testing days due to possible alterations in motor control related to LBP. Future studies should establish the measurement error in other subject populations, as the measurement error reported in this study may not necessarily generalize to clinical situations. This single-case feasibility study showed that a novice rater was not consistent across a broad range of measurement conditions, and future studies should examine if this pattern is generalizable to a larger group of novice raters.

**CONCLUSION**

A physical therapist, newly trained in RUSI, showed variable reliability when assessing muscles of the anterolateral abdominal wall during the abdominal drawing-in maneuver. While the novice assessor showed high reliability when measuring from stored images, the pattern of reliability was not consistent across other measurement factors, including the muscle measured and the state (rest and contracted) of the muscle for measurements conducted across images and across days. Inconsistencies in the pattern of results suggest that for a novice assessor using RUSI, training should be performed and reliability assessed for each abdominal muscle and measurement condition intended to be used for research and clinical practice.

**ACKNOWLEDGMENTS**

The authors would like to thank the subjects studied, Tom Johnsen (Scientific Officer, Division of Physiotherapy, The University of Queensland), Ross Darnell for statistical advice, and Mr Ian Wong for assistance with data collection.

**REFERENCES**


