Rehabilitative ultrasound imaging is a valid measure of trunk muscle size and activation during most isometric sub-maximal contractions: a systematic review

Shane L Koppenhaver[1], Jeffrey J Hebert[1], Eric C Parent[2] and Julie M Fritz[1]

[1]The University of Utah, USA, [2]The University of Alberta, Canada

Questions: Is rehabilitative ultrasound imaging a valid measure of trunk muscle size and activation? Are rehabilitative ultrasound imaging measures sensitive to change? **Design:** Systematic review of studies of criterion-related validity, construct validity, and sensitivity to change. **Participants:** People with low back pain and asymptomatic controls. **Outcome measure:** Trunk muscle size and activation measured by rehabilitative ultrasound imaging, MRI and/or EMG. **Results:** 37 studies were included. 10 studies investigated criterion-related validity and provided evidence that while ultrasound may be a valid measure of trunk muscle size, the validity of ultrasound to quantify muscle activation is context-dependent, depending on the muscle involved, the contraction strategy utilised, and the intensity of muscle contraction. 23 studies provided evidence of construct validity by demonstrating the ability of ultrasound measurement to differentiate individuals in terms of back pain, anthropometry, and postures. Six studies contained a limited amount of information about sensitivity to change. **Conclusions.** It is valid to use rehabilitative ultrasound imaging to measure trunk muscle size and activation during most isometric sub-maximal contractions. Ultrasound measures appear sensitive to both positive and negative change. **[Koppenhaver SL, Hebert JJ, Parent EC, Fritz JM (2009) Rehabilitative ultrasound imaging is a valid measure of trunk muscle size and activation during most isometric sub-maximal contractions: a systematic review. *Australian Journal of Physiotherapy* 55: 153–169]**

Key words: Systematic review, Validity of results, Ultrasonography, Rehabilitation, Skeletal muscle, Low back pain, Abdomen, Back, Physiotherapy

Introduction

For decades, ultrasound imaging has been used to assist with medical diagnosis by identifying structural tissue pathology. More recently, ultrasound imaging was introduced in rehabilitation to evaluate muscle morphology and function in persons with neuromusculoskeletal disorders such as low back pain (Teyhen 2007). In May 2006, an international panel of experts proposed a research agenda of rehabilitative ultrasound and adopted the term 'rehabilitative ultrasound imaging' to define the procedure of evaluating 'muscle and related soft tissue morphology and function during exercise and physical tasks' (Teyhen 2006). Muscle morphology refers to the shape, size, and structure of a muscle and may be important in rehabilitation as an indication of muscle atrophy and/or hypertrophy. With regard to rehabilitative ultrasound imaging, muscle morphology has, to date, generally focused on muscle size (eg, thickness and/or cross-sectional area) while muscle function has focused on the level and timing of muscle activation (Hodges 2005).

Rehabilitative ultrasound imaging is a relatively inexpensive, non-invasive, and safe imaging modality. These qualities make it an attractive method to quantify muscle size and activation. To date, most ultrasound research has focused on *transversus abdominis* and *lumbar multifidus* muscles, because dysfunction of these muscles has been linked to low back pain (Hodges and Richardson 1998, Hodges and Richardson 1999, Hodges 1999, Hodges 2001, Hungerford et al 2003, Yoshihara et al 2001, Zhao et al 2000). If ultrasound measures muscle size and activation accurately, it should be useful in research and clinical practice.

The ability of rehabilitative ultrasound imaging to quantify muscle size and activation depends on its reliability and validity. Using ultrasound to quantify muscle size is relatively straightforward and relies on its ability to measure muscle thickness and cross-sectional area. The level of muscle activation is then determined by comparing the size of a contracted muscle to its size during rest. Using measures of muscle size from static ultrasound images as an indication of muscle activation, however, this presents challenges. The level of muscle activation depends not just on a muscle's size, but on initial muscle (fascicle) length, amount of tendon stretch, type of contraction (isometric, concentric, or eccentric), muscle fibre pennation angle, and forces from surrounding tissues (Boyett et al 1991, Herbert et al 2002, Ito et al 1998, Narici et al 1996, Otten 1988, Takemori 1990). It is therefore necessary to examine the reliability and validity of ultrasound for quantifying muscle activation before its widespread use in research or clinical practice can be advocated. Rehabilitative ultrasound imaging has been found to have good within- and between-rater reliability (ICC 0.62 to 1.00), especially when used by experienced examiners and/or when multiple measurements are averaged (Hebert et al *in press*). Therefore, the specific research questions for this systematic review were:

1. Is rehabilitative ultrasound imaging a valid (criterion and construct) measure of trunk muscle size and activation?
2. Are rehabilitative ultrasound imaging measures sensitive to change?

Method

Identification and selection of studies

The search strategy used three important concepts from our research question: ultrasound, specific muscles, and the trunk region (see Appendix 1 on the eAddenda for full search strategy). The search was restricted to articles published in English or French language, as those were the only languages spoken by the review team. Ten databases were searched in May 2009 by two reviewers (JH and SK) independently. Titles and abstracts were screened, followed by the evaluation of full-text manuscripts against predetermined criteria (Box 1). The references of all included studies were searched to identify additional studies missed by our search strategy. The reviewers were blinded to each other's selections during each stage, but were not blinded to the studies' authors, journals, or results during the process. Disagreement by the two reviewers on study inclusion was resolved by discussion with a third reviewer (EP).

Box 1. Inclusion criteria.

- Published as full article, thesis or peer-reviewed report published in English of French.
- Included imaging of abdominal or lumbar muscles
- Reported on psychometric properties of quantitative measures of trunk muscles using ultrasound by comparing the measurements with repeated measurements (reliability), external measurements (concurrent validity), or longitudinal measurements (sensitivity to change).

Assessment of characteristics of the studies

Construct validity: The 2006 international ultrasound symposium described the theoretical constructs that ultrasound attempts to measure as 'muscle and related soft tissue morphology and function' (Teyhen 2006). Studies were classified as providing evidence of construct validity when quantitative ultrasound measurements were (1) able to distinguish between conditions or states known to be different on such constructs (known groups validity), or (2) compared either to an external measurement that is thought to reflect a similar construct and yielded similar results (convergent validity) or to a dissimilar construct and yielded dissimilar results (divergent validity) (Portney and Watkins 2008). For known groups validity, we considered 'known differences' to be present when published evidence of such differences exists or strong theoretical hypotheses would support such differences.

Criterion-related validity: Studies were classified as providing evidence of criterion-related validity when quantitative ultrasound measurements were compared to magnetic resonance imaging (MRI) or computerised tomography for measuring muscle size (Bemben 2002), or to electromyography (EMG) for measuring muscle activation (Price et al 2003).

Sensitivity to change: Studies were classified as providing evidence of sensitivity to change when quantitative ultrasound measurements were compared across a period of time in which a change in trunk muscle morphology or function was expected. Reviewers sought sensitivity to change indices if provided (eg, effect size, standardised response mean, Guyatt's responsiveness index) (Stratford et al 1996, Streiner and Norman 2003), but their inclusion was not necessary for study selection.

Quality: There is no standard list of criteria to review the quality of studies investigating validity and/or sensitivity to change. The Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool was recently developed to assess the quality of primary studies of diagnostic accuracy (Whiting et al 2003), has been evaluated (Hollingworth et al 2006, Whiting et al 2006), and is currently being used in the Cochrane Handbook for Diagnostic Test Accuracy Reviews (Smidt et al 2008). The 14 items in the QUADAS tool are each scored as 'yes', 'no', or 'unclear' (see Appendix 2 on the eAddenda for full explanation). The authors of QUADAS recommend avoiding summary scores because the potential biases associated with the items depend on the context in which they are applied (Whiting et al 2003). However, like other authors (Cook and Hegedus 2007) we included a '% yes' for each study to give readers a better understanding of overall study quality.

Additionally, the authors of QUADAS recognised that some items may need to be added or removed depending on the specific use of the tool (Whiting et al 2003). Both criterion-related validity and diagnostic accuracy are concerned with evaluating the relationship between a test measure and a criterion standard measure. Therefore, when using the QUADAS tool to evaluate the quality of criterion-related validity studies, we removed three items (Questions 10–12). When using the QUADAS tool to evaluate the quality of construct validity and sensitivity to change studies, we used only five items (Questions 1, 2, 8, 13, and 14) since raw quantitative measures from both ultrasound and from comparison reference standard tests (MRI or EMG) are arguably not subject to examiner interpretation.

Three reviewers were trained in applying the quality criteria. Two sample articles were assessed and the scores discussed. All articles included in this review were then independently assessed by two reviewers (SK and JH) and discrepancies between reviewers were resolved by discussion with a third reviewer (EP).

Data analysis

Data were extracted by two reviewers, independently, using customised forms. Publication details, contextual information, ultrasound equipment, ultrasound assessment procedures, quantitative measures, validity, and sensitivity to change were extracted. Results from data extraction were consolidated from the two reviewers and, if necessary, after discussion with a third reviewer.

All studies were included for analysis regardless of study quality. Studies within each classification (criterion-related, construct, and sensitivity to change) were qualitatively analysed and grouped by common themes. Criteria for interpreting the outcomes of studies were used as recommended by Terwee et al (2007). Studies were interpreted as supporting criterion-related validity if they reported a correlation and/or regression coefficient of ≥ 0.70. Studies were interpreted as supporting construct validity and sensitivity to change when they formulated specific hypotheses and at least 75% of the results were in accordance with those hypotheses.

Due to the heterogeneity of outcomes and methods of the included studies, meta-analysis was not performed.

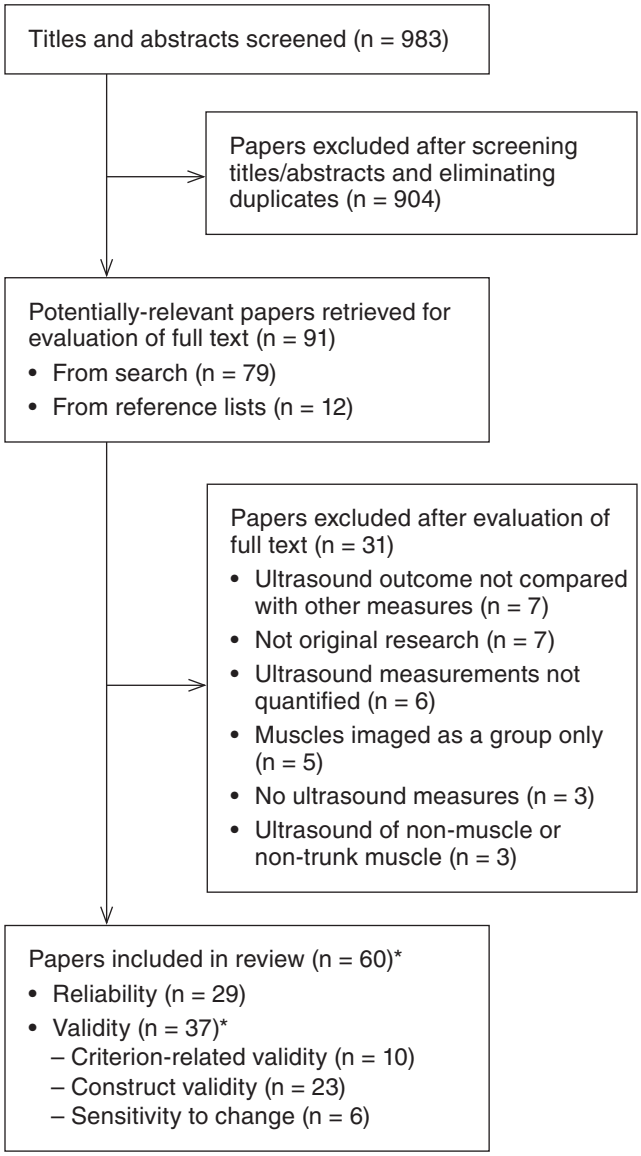

Figure 1. Flow of studies through review. * = papers may cover both reliability and validity, and more than one aspect of validity.

Results

Flow of studies through the review

Our search yielded 983 papers. After screening titles and abstracts and eliminating duplicates, 904 papers were excluded and 79 papers remained for detailed analysis (Figure 1). Evaluating the full text against the inclusion criteria, 31 of these were eliminated. From manually searching the reference lists of the 48 remaining papers, a further 12 papers were added. Of the 60 papers included, 10 contained information on criterion-related validity, 23 contained information on construct validity, and six contained information on the sensitivity to change of ultrasound. Given that some papers investigated multiple aspects of validity and/or sensitivity to change, 37 studies were reviewed and are summarised in Table 1.

Quality

The quality of included studies is presented in Table 2 for studies investigating criterion-related validity and Table 3 for studies investigating construct validity or sensitivity to change. Overall, the quality of the criterion-related validity studies was good, with more than 60% of the questions scoring 'yes' for all studies. The most consistent shortfalls were small sample sizes (mean n = 10 participants) and lack of representativeness of the participants compared with those who normally receive rehabilitative ultrasound imaging in clinical practice (primarily individuals with low back pain). However, only one of the 10 studies examining the criterion validity of ultrasound imaging included symptomatic patients.

The overall quality of the studies investigating construct validity and sensitivity to change was lower than those investigating criterion-related validity. Reviewers answered 'yes' to fewer than 50% of the questions for 12 of 28 studies. As with the criterion-related validity studies, the most consistent shortfall was that participants were not representative of those who receive rehabilitative ultrasound imaging in clinical practice. Only 11 of the 28 studies containing information regarding construct validity and sensitivity to change included patients with current low back pain. In the majority of these studies it was also unclear whether they had uninterpretable images and/or outcomes.

Criterion-related validity

Ten studies investigated criterion-related validity of rehabilitative ultrasound imaging (Table 4). Two studies investigated the ability of ultrasound to measure muscle size compared with MRI. One study examining *transversus abdominis* and *internal oblique* (Hides et al 2006) found substantial agreement (Shrout 1998) between ultrasound and MRI (ICC = 0.84 to 0.95). Similarly, another study examining *lumbar multifidus* (Hides et al 1995) found very small differences (0.03 to 0.21 cm^2) between ultrasound and MRI.

Five studies (Ferreira et al 2004, Hodges et al 2003, John and Beith 2007, Kiesel et al 2007a, McMeeken et al 2004) investigated the ability of ultrasound to measure muscle activation compared with EMG. One study (Kiesel et al 2007a) examining *lumbar multifidus* reported a strong linear relationship (r = 0.79) between ultrasound measures of thickness and EMG activity during 19–34% maximal voluntary contractions. Similarly, another study (McMeeken et al 2004) examining *transversus abdominis* reported a strong linear relationship between the change in muscle thickness and EMG activity ($R^2 = 0.87$) during contractions up to 80% maximal voluntary contraction. A third study (John and Beith 2007) examining *external oblique* found significant linear and curvilinear relationships between the change in muscle thickness and EMG activity in the majority of participants during isometric trunk rotation, but not during abdominal drawing-in. Of two studies examining *transversus abdominis, internal oblique,* and *external oblique* (Ferreira et al 2004, Hodges et al 2003), Hodges et al (2003) found the relationship between change in muscle thickness and EMG activity to be curvilinear for *transversus abdominis* and *internal oblique*, with muscle thickness increasing (almost linearly) only during contractions up to approximately 20% maximal voluntary contraction (r = 0.84 to 0.90), but little relationship during contraction of *external oblique* regardless of the strength of contraction (r = 0.23).

Instead of using B-mode ultrasound to measure muscle thickness or cross-sectional area, three of the most recent criterion-related studies used M-mode ultrasound and/or tissue doppler imaging to investigate the ability of ultrasound

Table 1. Summary of included validiy studies (n = 37).

Study	Participants	Body position	Muscle state/ contraction strategy	Criterion-related validity	Construct validity	Sensitivity to change
Akbari et al (2008)	n = 49 (Chronic LBP) Age *(yr)* = 40.0 (SD 3.6) Gender = not reported	Prone (LM) Supine (TrA)	Rest only			Change in resting LM and TrA thickness following 8 weeks of either a motor control or general trunk exercise program
Coldron et al (2003)	n = 20 (asymptomatic) Age *(yr)* = not reported Gender = 20 F	Side-lying and prone	Rest only		Convergence of LM CSA side lying and LM CSA prone	
Critchley et al (2002a)	n = 20 (asymptomatic) Age *(yr)* = not reported Gender = 12 F, 8 M	Four-point kneeling	Abdominal hollowing with and without pelvic floor contraction		Different abdominal muscle activation with pelvic floor contraction than without	
Critchley et al (2002b)	n = 42 (mixed) Age *(yr)* = LBP: 40.1 (SD 10.8) Control: 32.2 (SD 11.3) Gender = 23 F, 19 M	Four-point kneeling	Abdominal hollowing		Different increase in TrA thickness in patients with LBP than controls	
Endleman et al (2008)	n = 26 (asymptomatic) Age *(yr)* = 43 (SD 14) Gender = 18 F, 8 M	Supine	Various pilates exercises		Different TrA and IO thickness between rest, incorrectly performed pilates, and correctly performed pilates exercises	
Ferreira et al (2004)	n = 20 (mixed) Age *(yr)* = LBP: 27.8 (SD 5.1) Control: 32.7 (SD 10.6) Gender = not reported	Supine	Isometric leg flexion and extension	Test = Changes in TrA, IO, and EO thickness Criterion = RMS EMG amplitude	Change of TrA, IO, and EO thickness during leg flexion and extension. Difference in change between patients with LBP and controls	
Hides et al (1994)	n = 77 (mixed) Age *(yr)* = LBP: 31 (SD 8) Control: 25 (SD 4) Gender = 40 F, 37 M	Prone	Rest only		Different LM CSA between painful and non-painful sides and levels	
Hides et al (1995)	n = 10 (asymptomatic) Age *(yr)* = not reported Gender = 10 F	Prone	Rest only	Test = LM CSA Criterion = MRI		
Hides et al (1996 & 2001)	n = 39 (acute LBP) Age *(yr)* = 30.7 (SD 8) Gender = 23 F, 16 M	Prone	Rest only		Different 4 & 10 wk LM CSA asymmetry in those experiencing high LBP recurrence rates at 1 and 2–3 years	Change in LM CSA asymmetry after 4 and10 weeks in patients treated with specific stabilisation exercises.

Study	Participants	Body position	Muscle state/ contraction strategy	Criterion-related validity	Construct validity	Sensitivity to change
Hides et al (2006)	n = 13 (asymptomatic) Age *(yr)* = 21.3 (SD 2.1) Gender = 13 M	Supine	Active drawing-in manoeuvre	Test = TrA and OI thickness, and TrA fascial slide Criterion = MRI		
Hides et al (2007)	n = 19 (asymptomatic) Age *(yr)* = 20.3 (SD 5.0) Gender = 11 F, 8M	Supine	Unilateral isometric leg extension		Different TrA thickness between rest and simulated weight bearing	
Hides et al (2008)	n = 26 (mixed) Age *(yr)* = 21.7 (SD 2.3) Gender = 26 M	Prone	Rest only		Different LM CSA asymmetry in those with LBP than those without	Change in LM CSA asymmetry after 13 weeks of cricket training, with or without specific stabilization exercises
Hides et al (2009)	n = 39 (mixed) Age *(yr)* = LBP: 28.1 (SD 10.3) Control: 24.4 (SD 5.7) Gender = 17 F, 22 M	Supine	Unilateral isometric leg extension		Different TrA shortening in those with LBP than controls	
Hodges et al (2003)	n = 3 (asymptomatic) Age *(yr)* = not reported Gender = 3 M	Sitting	Isometric contraction	Test = Changes in TrA, IO, and EO thickness Criterion = RMS EMG amplitude		
Hodges et al (2006)	n = 21 (porcine) Age *(yr)* = 0.3 Gender = not reported	not reported	Rest only		Not included	Change in LM CSA 3 & 6 days following induced injury.
Ishida et al (1994)	n = 92 (asymptomatic) Age *(yr)* = 25 (SD 2) Gender = 92 F	not reported	Rest only		Different abdominal muscle thickness between Caucasians and Japanese.	
John et al (2007)	n = 24 (asymptomatic) Age *(yr)* = 24.5 (SEM 0.5) Gender = 15 F, 9 M	Supine	Isometric trunk rotation and lower abdominal drawing-in	Test = Changes in EO thickness Criterion = Surface EMG		
Keisel et al (2007a)	n = 5 (asymptomatic) Age *(yr)* = 28.0 (SD 5.6) Gender = 3 F, 2 M	Prone	Contralateral arm raise	Test = Changes in LM thickness Criterion = RMS EMG amplitude		
Keisel et al (2007b)	n = 76 (mixed) Age *(yr)* = 43.1 (SD 10.9) Gender = 35 F, 21 M	Prone (LM) Supine (TrA)	Contralateral arm raise (LM) and abdominal drawing-in (TrA)		Different increase in TrA and LM thickness between patients with LBP and controls	
Kiesel et al (2008)	n = 6 (asymptomatic) Age *(yr)* = 26.0 (SD 7.3) Gender = 6 M	Prone (LM) Supine (TrA)	Contralateral arm raise (LM) & abdominal drawing-in (TrA)			Change in contracted LM and TrA thickness following induced LM muscle pain.

Table 1. Summary of included validiy studies (n = 37) – *continued*.

Study	Participants	Body position	Muscle state/ contraction strategy	Criterion-related validity	Construct validity	Sensitivity to change
Lee et al (2006)	n = 35 (mixed) Age *(yr)* = LBP: 39.9 Control: 41.7 Gender = 35 M	Prone Standing	Rest only		Different LM CSA between patients with LBP and controls during prone lying, upright standing, and 25° and 45° forward stooping.	
Mannion et al (2008)	n = 14 (asymptomatic) Age *(yr)* = Men: 22.3 (SD 1.5) Women: 24.0 (SD 2.3) Gender = 6 F, 8 M	Standing	Contralateral shoulder flexion, abduction, and extension	Test = Onset of TrA, IO, and EO activity Criterion = EMG		
Masuda et al (2005)	n = 50 (asymptomatic) Age ((*yr)* = 30.4 (SD 6.6) Gender = 50 M	Standing	Rest only		Convergence of posture-related change of ES muscle thickness with ES muscle oxygenation and blood volume changes	
McMeeken et al (2004)	n = 9 (asymptomatic) Age *(yr)* = 40.7 (SEM 2.7) Gender = 5 F, 4 M	Supine	Abdominal hollowing	Test = Changes in TrA thickness Criterion = RMS EMG amplitude		
Misuri et al (1997)	n = 6 (asymptomatic) Age *(yr)* = not reported Gender = 6 M	Sitting	Rest only		Different TrA, IO, EO, and RA thickness during phases of respiration	
Oguri et al (2004)	n = 28 (asymptomatic) Age *(yr)* = 62.1 (SD 2.8) Gender = 28 M	Prone Supine	Rest only		Different trunk muscle thickness between high and intermediate/low level runners. Convergence of trunk muscle thickness measures and weekly training distances	
Rankin et al (2006)	n = 123 (asymptomatic) Age ((*yr)* = 40.6 (SD 14.1) Gender = 68F, 55M	Supine	Rest only		Different muscle thicknesses between men and women and based on age	
Reeve et al (2009)	n = 20 (asymptomatic) Age *(yr)* = 29 Gender = 10 F, 10 M	Supine Erect sitting Slouched sitting Erect standing Sway-backed standing	Rest only		Different TrA thickness between different postures	

Study	Participants	Body position	Muscle state/ contraction strategy	Criterion-related validity	Construct validity	Sensitivity to change
Springer et al (2006)	n = 32 (asymptomatic) Age *(yr)* = 31.9 (SD 7.8) Gender = 15 F, 17 M	Supine	Abdominal drawing-in manoeuvre		Different trunk muscle thickness between men and women. Trunk muscle thickness measures converge with BMI	
Stanton et al (2008)	n = 28 (asymptomatic) Age *(yr)* = Men: 28.1 (SD 7.42) Women: 26.5 (SD 5.49) Gender = 14 F, 14 M	Prone	Abdominal hollowing Abdominal brace		Different TrA CSA between rest, abdominal hollowing, and abdominal brace	
Stokes et al (2005)	n = 120 (asymptomatic) Age *(yr)* = Female: 34.2 (SD 12.8) Male: 40.1 (SD 13.0) Gender = 68 F, 52 M	Prone	Rest only		Different LM muscle thickness between men and women and between levels. Muscle thickness measures converge with CSA measures	
Van et al (2006)	n = 25 (asymptomatic) Age *(yr)* = 19.5 (SD 2.2) Gender = 19 F, 6 M	Prone	Volitional contraction ('swell the LM')			Change in amount of LM contraction during training session and after 1 week of retention
Vasseljen et al (2006)	n = 10 (asymptomatic) Age *(yr)* = 28.4 (F), 26.8 (M) Gender = 5 F, 5 M	Standing	Ipsilateral arm raise	Test = Onset of LM activity Criterion = EMG		
Vasseljen et al (2009)	n = 10 (asymptomatic) Age *(yr)* = not reported Gender = 7 F, 3 M	Standing	Ipsilateral arm raise	Test = Onset of TrA, IO, and EO activity Criterion = EMG		
Wallwork et al (2008)	n = 34 (mixed) Age *(yr)* = LBP: 41.9 (SD 13.7) Control: 33.9 (SD 11.2) Gender = 18 F, 16 M	Prone	CSA at rest Thickness change during volitional contraction ('swell the LM')		Different LM CSA asymmetry and change in LM thickness at L5 between those with LBP and those without	
Whittaker et al (2008)	n = 24 (LBP) Age (*(yr)* = 43 (SD 10) Gender = 21 F, 3 M	Supine	Respiration		Different TrA thickness change between patients with LBP and patients with LBP + hypocapnia	

F = female, M = male, TrA = transversus abdominis, IO = internal oblique, EO = external oblique, LM = lumbar multifidus, MRI = magnetic resonance imaging, RMS = root mean square, EMG = electromyography, CSA = cross sectional area, LBP = low back pain

Table 2. QUADAS score for studies investigating criterion-related validity (n = 10).

Study	Participants representative of clinical practice	Selection criteria described	Classification of target condition by reference standard	Short time between reference standard and index test	Verification of sample using reference standard	Same reference standard regardless of index test result	Reference standard independent of index test	Execution of index test described	Execution of reference standard described	Uninterpretable results reported	Withdrawals explained	% Yes
Ferreira et al (2004)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	100
Hides et al (2006)	N	Y	Y	Y	Y	Y	Y	Y	Y	Unclear	Unclear	73
Hides et al (1995)	N	Y	Y	Unclear	Y	Y	Y	Y	Y	Y	Y	82
Hodges et al (2003)	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	82
John et al (2007)	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	91
Kiesel et al (2008)	N	Y	Y	Y	Y	Y	Y	Y	Y	Unclear	Unclear	73
Mannion et al (2008)	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	91
McMeeken et al (2004)	N	N	Y	Y	Y	Y	Y	Y	Y	Unclear	Unclear	64
Vasseljen et al (2006)	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Unclear	73
Vasseljen et al (2009)	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	91

QUADAS = Quality Assessment of Diagnostic Accuracy Studies

to measure the onset of muscle activity compared with EMG (Mannion et al 2008, Vasseljen et al 2006, Vasseljen et al. 2009). In *lumbar multifidus*, Vasseljen et al (2006) found good agreement between M-mode ultrasound and EMG (ICC = 0.89 to 0.93) even though ultrasound measurements of muscle onset were an average of 16 ms (95% CI 11 to 23) slower than concurrent EMG measurements. Using tissue doppler imaging of *transversus abdominis, internal oblique,* and *external oblique*, Mannion et al (2008) found a moderate correlation (r = 0.47) and a 20 ms (SD 30) systematic delay of ultrasound measurement of earliest muscle onset when compared to EMG. Similarly, using both M-mode ultrasound and tissue doppler imaging of *transversus abdominis, internal oblique,* and *external oblique*, Vasseljen et al (2009) found a systematic delay of muscle onset with both techniques compared with EMG ranging from 2 to 15 ms, except for *external oblique* which contracted 54 ms earlier according to ultrasound compared with EMG.

Construct validity

Twenty-three studies investigated the construct validity of rehabilitative ultrasound imaging (Table 5). Eight studies provided evidence of 'known groups' validity by demonstrating different ultrasound measurements between groups with and without pain. Of these eight studies, four found significantly attenuated *transversus abdominis* activation (thickness change and/or 'slide') in patients with low back pain compared with asymptomatic controls (Critchley and Coutts 2002, Ferreira et al 2004, Hides et al 2009, Kiesel et al 2007b). Five studies found differences in *lumbar multifidus* such as more side-to-side asymmetry in cross-sectional area (Hides et al 1994, Hides et al 2008, Wallwork et al 2008), attenuated activation (thickness change) (Kiesel et al 2007b, Wallwork et al 2008), and a different posture-related pattern of cross-sectional area (Lee et al 2006) between patients with low back pain compared with asymptomatic controls. The one study investigating *internal oblique* and *external oblique* found no significant difference in activation in patients with low back pain compared with asymptomatic controls (Ferreira et al 2004). The final pain-related study found that a group who undertook specific stabilisation exercises for 4 weeks had less side-to-side asymmetry in *lumbar multifidus* and less recurrence of low back pain in the following 2–3 years than a group managed medically (Hides et al 2001).

Four studies provided evidence of 'known groups' validity by demonstrating different trunk muscle morphology between groups differing in terms of sex, race, or age (Ishida et al 1994, Rankin et al 2006, Springer et al 2006, Stokes et al 2005). Eight additional studies provided evidence of 'known groups' validity by demonstrating different trunk muscle size and/or activation between different postures (Lee et al 2006, Reeve & Dilley 2009), between different activities (Critchley 2002, Endleman & Critchley 2008, Ferreira et al 2004, Hides et al 2007, Misuri et al 1997, Stanton & Kawchuk 2008), between groups with and without breathing disorders (Whittaker 2008), and between different levels of activity (Oguri et al 2004). Finally, five studies provided evidence of convergent validity by demonstrating the convergence of ultrasound measurements with measurements of muscle oxygenation and change in blood volume (Masuda et al 2005), activity level (Oguri et al 2004), body mass index (Springer et al 2006), and other measures of the same muscle (Coldron et al 2003, Stokes et al 2005).

Table 3. QUADAS score for studies investigating construct validity and sensitivity to change (n = 28).

Study	Participants representative of clinical practice	Selection criteria described	Execution of index test described	Uninterpretable results reported	Withdrawals explained	% Yes
Akbari et al (2008)	Y	Y	Unclear	Y	Y	80
Coldron et al (2003)	N	Y	Y	Y	Y	80
Critchley et al (2002a)	Y	Y	Y	Unclear	Y	80
Critchley et al (2002b)	N	Y	Y	Y	Y	80
Endleman et al (2008)	Y	Y	Y	Y	Y	100
Ferreira et al (2004)	Y	Y	Y	Y	Y	100
Hides et al (1994)	Y	Y	Y	Y	Y	100
Hides et al (1996)	N	Y	N	Unclear	Y	40
Hides et al (2001)	N	Y	N	Unclear	Unclear	20
Hides et al (2007)	N	Y	Y	Y	Y	80
Hides et al (2008)	Y	N	Y	Y	Y	80
Hides et al (2009)	Y	Y	Y	Y	Y	100
Hodges et al (2006)	N	N	Y	Unclear	Y	40
Ishida et al (1994)	N	N	N	Unclear	Y	20
Kiesel et al (2007a)	N	N	Y	Unclear	Y	40
Keisel et al (2007b)	Y	Y	Y	Y	Y	100
Lee et al (2006)	N	Y	N	Unclear	Y	40
Masuda et al (2005)	N	N	Y	Unclear	Y	40
Misuri et al (1997)	N	N	Y	Unclear	Y	40
Oguri et al (2004)	Y	Y	N	Unclear	Y	60
Rankin et al (2006)	N	N	Y	Unclear	Unclear	20
Reeve et al (2009)	N	Y	Y	Y	Y	80
Springer et al (2006)	N	N	Y	Unclear	Y	40
Stanton et al (2008)	N	Y	Y	Y	Y	80
Stokes et al (2005)	N	N	Y	Unclear	Y	40
Van et al (2006)	N	Y	Y	Unclear	Unclear	40
Wallwork et al (2008)	Y	Y	Y	Unclear	Unclear	60
Whittaker et al (2008)	Y	Y	Y	Y	Y	100

QUADAS = Quality Assessment of Diagnostic Accuracy Studies

Sensitivity to change

Six studies provided evidence of rehabilitative ultrasound imaging's sensitivity to change. Although none of these studies reported indices of sensitivity to change such as effect size, standardised response mean, or Guyatt's responsiveness index, three provided enough information to calculate such an index (Table 6).

Four of the six studies demonstrated increased muscle size and/or activation following training (Akbari et al 2008, Hides et al 1996, Hides et al 2008, Van et al 2006). Akbari et al (2008) reported an increase in resting thickness in both *transversus abdominis* and *lumbar multifidus* after eight weeks of motor control exercises or general exercises in patients with chronic low back pain. The authors reported Cohen's effect sizes (Stratford et al 1996, Streiner and Norman 2003) of 0.83 (motor control exercise) versus 0.59 (general exercise) for *transversus abdominis*, and 0.45 (motor control exercise) versus 0.28 (general exercise) for *lumbar multifidus*. Hides et al (2008) reported less side-to-side asymmetry in *lumbar multifidus* cross-sectional area after 13 weeks of intense cricket training camp for elite cricketers with and without low back pain. The largest increase in cross-sectional area occurred at the L5 level on the smaller side. The authors reported Cohen's effect sizes of 1.10 (for those with back pain who also undertook stabilisation exercises) versus 0.23 (for those without back pain who did not undertake the exercises). Hides et al (1996) also reported less side-to-side asymmetry in *lumbar multifidus* cross-sectional area at 4 and 10 weeks in patients with acute low back pain compared with patients that were medically managed, even though pain and disability had largely resolved by 4 weeks. Van et al (2006) reported an increase in ability to contract *lumbar multifidus* over a single session in both a group that received visual ultrasound biofeedback and a group that received verbal biofeedback. However, only the group receiving ultrasound visual biofeedback retained the increase after one week.

Two of the six studies demonstrated decreased muscle size after inducing pain and/or injury (Hodges et al 2006, Kiesel et al 2008). Hodges et al (2006) reported a decrease in *lumbar multifidus* cross-sectional area three and six days after injury to the L3/4 intervertebral disc and L4 nerve roots in pigs. After inducing pain by injecting hypertonic saline into the *longissimus* muscle at the L4 level, Kiesel et al (2008) reported a decrease in muscle thickness during contraction of 0.09 cm (Cohen's effect size of 1.13) for *transversus abdominis* and of 0.26–0.32 cm (Cohen's effect size 0.44–0.53) for *lumbar multifidus*.

Table 4. Criterion-related validity of rehabilitative ultrasound imaging of muscles of the trunk.

Study	Results	Estimates	Evidence of criterion-related validity
Muscle size compared to MRI			
Hides et al (2006)	High level of agreement between RUSI and MRI measurements	ICC TrA = 0.84 to 0.94 IO = 0.91 to 0.95 TrA slide = 0.78 to 0.91	Yes
Hides et al (1995)	No significant difference between US and MRI measurements at any level	*p* value L2 to S1 = > 0.05 Mean difference L5 = 0.03 cm^2 S1 = 0.21 cm^2	Yes
Level of activation compared to EMG			
Ferreira et al (2004)	RUSI and EMG measurements 'agreed', but level of agreement not analysed statistically	None	Partial
Hodges et al (2003)	Curvilinear relationships between RUSI and EMG measures for the TrA and IO, but not EO	r TrA = 0.90, $p < 0.005$ IO = 0.84, $p < 0.005$ EO = 0.23, $p = 0.413$	Partial
John et al (2007)	Significant curvilinear and linear relationship between RUSI and EMG with contractions up to 80% MVC for the majority of participants (21/24) during isometric trunk rotation, but not during abdominal drawing-in. Data not pooled due to high inter-participant variability.	R^2 Isometric trunk rotation = 0.60 to 0.80 for majority Abdominal drawing-in = –0.51 to 0.73	Partial
Keisel et al (2007)	LM muscle thickness change was highly correlated with EMG activity between 19% and 34% MVC	r = 0.79, $p < 0.001$	Yes
McMeeken et al (2004)	Linear relationships between RUSI and EMG measures with contractions up to 80% MVC for the TrA.	$R^2 = 0.87$, $p < 0.001$	Yes
Timing of activation compared to EMG			
Mannion et al (2008)	Significant correlation between RUSI (TDI) and EMG, but RUSI measured earliest muscle onset consistently later than EMG	r = 0.47, p < 0.0001 Limits of agreement = ± 60 ms Mean difference = 20 ms	Partial
Vasseljen et al (2006)	Strong correlation between RUSI and EMG, but RUSI measured LM onset consistently later than EMG	ICC = 0.89 to 0.93 Limits of agreement = ± 43 ms Mean difference = 16 ms	Yes
Vasseljen et al (2009)	Except for M-mode of EO, both M-mode US and TDI US consistently measured muscle onset later than EMG for TrA, IO, and EO	Mean difference M-mode TrA = 7 ms M-mode IO = 2 ms M-mode EO = 54 ms (earlier) TDI TrA = 7 ms TDI IO = 15 ms TDI EO = 12 ms	Partial

TrA = transversus abdominis, IO = internal oblique, EO = external oblique, EMG = electromyography, LM = lumbar multifidus, MRI = magnetic resonance imaging, EMG = electromyography, MVC = maximal voluntary contraction, RUSI = realtime ultrasound imaging, US = ultrasound, TDI = tissue doppler imaging.

Discussion

This is the first study to systematically review the evidence of the validity of using rehabilitative ultrasound imaging to quantify the size and activation of trunk muscles. Our primary findings are (1) it is valid to use ultrasound to measure trunk muscle size and activation during most isometric, sub-maximal contractions, and (2) ultrasound measures of trunk muscle size and activation appear sensitive to both positive and negative change.

Criterion-related validity

Muscle size: The substantial agreement between ultrasound and MRI measurements of thickness and cross-sectional area provide evidence of the validity of ultrasound to measure trunk muscle size accurately (Hides et al 1996, 2006). Two studies investigated different muscles during different conditions and found similar results. Unfortunately, both studies used small samples of asymptomatic individuals, which may reduce the generalisability of the findings to those with low back pain. On the other hand, we are not

Table 5. Construct validity of rehabilitative ultrasound imaging of muscles of the trunk.

Study	Results	Estimate	Evidence of construct validity
Pain-related conditions			
Critchley et al (2002b)	Mean increase in TrA thickness significantly smaller in LBP patients than in controls	Mean LBP patients = 19% Controls = 50% $p < 0.001$	Yes
Ferreira et al (2004)	TrA and IO thickness change during isometric leg flexion and extension smaller in those with LBP than in controls for the TrA, but not the IO or EO	*p* value TrA < 0.01 IO = 0.31 EO = 0.85	Yes
Hides et al (1994)	Asymmetry in LM CSA larger in LBP patients than in asymptomatics. Asymmetry at the level of symptoms greater than at levels above and below	Mean LBP = 31% Asymptomatics = 3% *p* value Between groups < 0.001 Between levels < 0.05	Yes
Hides et al (2001)	Patients with larger LM asymmetry at 4 and 10 wks (managed medically) had higher recurrence of LBP at 1 and 2 to 3 years than those with minimal asymmetry at 4 and 10 weeks (treated with specific stabilisation exercises)	Mean Medically managed 1 yr = 84% Specific exercise 1 yr = 30% Medically managed 2–3 yr = 75% Specific exercise 2–3 yr = 35% $p < 0.01$	Yes
Hides et al (2008)	After controlling for age, height, and body mass, patients with LBP had more side-to-side asymmetry in LM CSA than those without LBP at every lumbar vertebral level	Mean LBP = 1.8 to 8.3% Controls = 1.2 to 5.7% $p < 0.05$	Yes
Hides et al (2009)	Patients with LBP had significantly less TrA shortening ($p < .0001$) than controls	Mean LBP = 2.4 mm (SD 3.1) (25% body weight) Controls = 4.3 mm (SD 1.4) (25% body weight) LBP = 3.6 mm (SD 3.6) (45% body weight) Controls = 6.7 mm (SD 1.9) (45% body weight)	Yes
Keisel et al (2007b)	Asymptomatic individuals had significantly more TrA thickness change than patients with LBP classified in the stabilisation, mobilisation, or specific exercise classification, significantly more LM thickness change at L4/5 than patients classified in either the stabilization or specific exercise classification, and significantly more LM thickness change at L5/S1 than patients classified in the specific exercise classification	Mean Control TrA = 99% LBP TrA = 52 to 67% Control LM = 13 to 24% LBP LM = 8 to 17% $p < 0.05$	Yes
Lee et al (2006)	In controls, LM CSA increased from prone lying to upright standing and then gradually decreased from 25° to 45° forward stooping. A reverse pattern of CSA change was identified in patients with chronic LBP.	Too numerous to summarise *p* value < 0.05	Yes
Wallwork et al (2008)	Patients with chronic LBP had significantly smaller LM CSA and less contracted thickness change at L5 than controls after adjustment for age, weight, and height	Mean difference CSA = 1.75 cm^2 Thickness = 3.24% $p < 0.05$	Yes
Position or activity-related conditions			
Coldron et al (2003)	High agreement between LM CSA side lying and LM CSA prone	r = 0.90 to 0.91 Mean difference = 0.03 to 0.15 cm Limits of agreement = ± 0.94 to 0.99 cm^2	Yes
Critchley et al (2002a)	TrA thickness change during low abdominal hollowing increased when pelvic floor contraction was added	Mean Without pelvic floor = 49.7% With pelvic floor = 65.8% $p = 0.015$	Yes

Endleman et al (2008)	TrA and IO thickness significantly larger during all correctly performed pilates exercises when compared to rest and to the incorrect imprint exercise.	Too numerous to summarise *p* values < 0.001	Yes
Hides et al (2007)	Both IO and TrA muscles significantly thickened during unilateral simulated weight-bearing task	Mean increase IO = 18.5% (SD 9.7) TrA = 24.7% (SD 17.5) *p* < 0.0001	Yes
Masuda et al (2005)	ES thickness, blood volume, and oxygenated haemoglobin simultaneously increased during relaxed extension and decreased during relaxed flexion	Too numerous to summarise *p* values < 0.05	Yes
Misuri et al (1997)	Thickness of the TrA, IO, and RA (but not EO) increased during maximal expiratory manoeuvres	Too numerous to summarise *p* values < 0.02	Yes
Oguri et al (2004)	ES thickness larger in high-level runners than both intermediate and low-level runners. Significant correlation between ES thickness and training distance	r = 0.75, *p* < 0.001 Mean difference High-level to intermediate = 15%, *p* < 0.01 High-level to low-level = 12%, *p* < 0.05	Yes
Reeve et al (2009)	TrA thickness significantly larger in erect standing than in sway-back standing and in erect sitting than in slouched sitting	Mean Erect standing = 4.6 mm (SD 1.4) Sway-back standing = 3.3 mm (SD 1) Erect sitting = 4.3 mm (SD 1.6) Slouched sitting = 3.5 mm (SD 1.1) *p* < 0.001	Yes
Stanton et al (2008)	TrA CSA significantly larger during both abdominal hollowing and abdominal brace than during rest	Mean CSA not reported *p* value < 0.01	Yes
Whittaker et al (2008)	TrA thickness change significantly larger during respiration in patients with LBP + hypocapnia than in patients with LBP only	Mean LBP + hypocapnia = 21% (SD 7.6) LBP only = 1% (SD 5.8) *p* < 0.001	Yes
Anthropometric conditions			
Ishida et al (1994)	Abdominal muscles significantly thicker in Caucasian than Japanese females	Mean Caucasian = 12.1 mm Japanese =10.0 mm *p* < 0.001	Yes
Rankin et al (2006)	Males had significantly larger TrA, IO, EO, and RA muscles than females. Muscle thicknesses significantly, but weakly correlated with age	Means = too numerous to summarise r = –0.27 to –0.04 *p* values < 0.001	Yes
Springer et al (2006)	Men had greater TrA and total lateral abdominal muscle thickness than women. BMI correlated with TrA thickness at rest and while contracted. No differences found based on hand dominance	Means = too numerous to summarise r Rest = 0.66 Contracted = 0.77 *p* < 0.01	Partial
Stokes et al (2005)	LM CSA larger in males than females and larger at L5 than L4. LM CSA correlated with LM thickness, spinous process length, and laminar width. LM CSA did not differ based on age	Mean Males L4 = 7.87 cm^2 Males L5 = 8.91cm^2 Females L4 = 5.55 cm^2 Females L5 = 6.65 cm^2 r LM thickness = 0.54 to 0.80 Spinous process length = 0.38 to 0.60 Laminar width = 0.36 to 0.52 *p* < 0.05	Yes

TrA = transversus abdominis, IO = internal oblique, EO = external oblique, LM = lumbar multifidus, BMI = body mass index, CSA = cross-sectional area, LBP = low back pain

Table 6. Sensitivity to change of rehabilitative ultrasound imaging of muscles of the trunk.

Study	Results	Estimates	Evidence of sensitivity to change
Increase with training			
Akbari et al (2008)	TrA and LM thickness significantly increased following 8 weeks of either a motor control or general trunk exercise program	Mean change TrA motor control = 0.5 mm (from 1.9 mm) TrA general exercise = 0.3 mm (from 1.9 mm) LM motor control = 1.1 mm (from 8.6 mm) LM general exercise = 0.4 mm (from 8.8 mm)	Yes
Hides et al (1996)	After 4 weeks less LM CSA asymmetry found in patients treated with specific stabilisation exercises than in those managed medically	Mean Exercise 4 wks = 0.7% Exercise 10 wks = 0.2% Management 4 wks = 16.8% Management 10 wks = 14.0% $p < 0.001$	Yes
Hides et al (2008)	LM CSA asymmetry significantly decreased over the 13-week training period in both groups and at all lumbar vertebral levels. Decreased asymmetry during training at the L5 level was greater for the LBP group receiving stabilisation training than the non-LBP group	Mean change Stabilisation = 8.3% to 1.4% Non-LBP = 0.8% to 0.5% $p < 0.05$	Yes
Van et al (2006)	Both groups improved ability to contract LM during the training session (mean changes not reported). RUSI visual biofeedback group retained the improved percent activation after 1 week, while the verbal biofeedback group regressed towards baseline (reported)	Mean change RUSI = 13% to 13% Verbal = 6% to 3% $p < 0.001$	Yes
Decrease with pain or injury			
Kiesel et al (2008)	After injection, contracted TrA and LM thickness decreased	Mean change TrA = 0.09 cm (from 0.68 to 0.59 cm) LM = 0.29 cm (from 4.09 to 3.80 cm) $p < 0.01$	Yes
Hodges et al (2006)	After injury to the L3/4 disc, LM CSA was reduced at L4 on the side of the lesion. After transection of the L3 nerve root, LM CSA was reduced at L4, L5, and L6. No change after sham injury or on the opposite side	Mean change Disc = 17% Nerve root L4 = 13% Nerve root L5 = 20% Nerve root L6 = 12% $p < 0.001$	Yes

LBP = low back pain, TrA = transversus abdominis, LM = lumbar multifidus, CSA = cross-sectional area RUSI = realtime ultrasound imaging

aware of any suggestion that ultrasound as a measure of muscle morphology is population- or condition-specific.

Level of muscle activation: Using rehabilitative ultrasound imaging to measure muscle activation is more complex and relies on the accuracy of changes in muscle size during contraction reflecting muscle activation. It is intuitive that muscles generally change size when they contract as muscle fibres usually shorten with the increase in overlap of actin and myosin filaments. The change in muscle size, however, also depends on the type of contraction (isometric, concentric, or eccentric) and fibre arrangement of the contracting muscle (parallel or pennate) (Boyett et al 1991, Herbert et al 2002, Ito et al 1998, Narici et al 1996, Otten 1988, Takemori 1990). If the length of a muscle is changed, then its size (thickness and/or cross-sectional area) will also change regardless of the level of activation. During isometric contraction, muscle thickening and shortening is dependent upon tendon stretch. At low contraction intensity, muscles can thicken and shorten more because the tendon stretches more whereas at higher intensity the muscle thickens and shortens less because tendon stretch lessens with increasing tendon length (Herbert et al. 2002). In pennate muscle, changes in muscle size may also depend upon changes in fibre pennation angle which becomes greater (more perpendicular to the tendon) during contraction. Furthermore, changes in muscle size during a contraction depend upon competing forces from surrounding tissues (Hodges 2005). For example, contracting abdominal muscles may compress each other, thereby reducing any associated increased thickness and/or cross-sectional area. Finally, making inferences about the level of muscle activation from changes in muscle size is dependent upon initial muscle activity. For changes in muscle size to represent the level of

muscle activation, the measured muscle must be initially at rest (electrically silent).

The results of studies comparing ultrasound measurements with EMG activity suggest that the ability of ultrasound to measure muscle activation is complex and probably context dependent. There is evidence that ultrasound is accurate at measuring muscle activation of *transversus abdominis*, *internal oblique*, *external oblique,* and *lumbar multifidus* (Ferreira et al 2004, Hodges et al 2003, John and Beith 2007, Kiesel et al 2007a, McMeeken et al 2004, Vasseljen et al 2006). However, the relationship between ultrasound measurements and EMG activity may depend upon both the intensity of muscle contraction and the contraction strategy used. The only study to investigate *lumbar multifidus* did so within a limited range (19–34% of maximal voluntary contraction) (Kiesel et al 2007a). Although the author found a linear relationship between the two measures, we do not know anything about the relationship outside this range. Two studies of *transversus abdominis* assessed a wider range of contraction intensities (Hodges et al 2003, McMeeken et al 2004). These studies provide evidence of the ability of ultrasound to measure *transversus abdominis* muscle activation at low levels of contraction. However, the studies had conflicting results regarding stronger contractions and came to different conclusions about the relationship between ultrasound and EMG measurements. One reported a linear relationship (McMeeken et al 2004) while the other reported a curvilinear relationship, finding approximate linearity only with contractions below 20% of maximal voluntary contraction (Hodges et al 2003). These studies used different contraction strategies, which may explain the difference in findings. McMeeken et al (2004) used an abdominal hollowing task that may allow more shortening of *transversus abdominis* than the isometric contraction used by Hodges et al (2003). The difference in findings could also be the result of inter-participant variability as both studies examined a small number of participants (n = 3 and 9). Hodges et al (2003) also studied *external oblique* and found that changes in muscle activation on EMG were not associated with changes in muscle thickness during an isometric abdominal contraction. This finding agreed with those of John et al (2007) who reported that ultrasound could measure *external oblique* muscle activation during isometric trunk rotation, but not during abdominal drawing-in. This suggests that the ability of ultrasound to measure muscle activation may depend not only upon the level of contraction, but also on the contraction strategy used. While the reason for this remains unknown, it may be partially explained by competing forces of surrounding muscles. Both *transversus abdominis* and *internal oblique* became thicker during the isometric contraction (Hodges et al 2003) and may have compressed *external oblique* thereby limiting its expansion. Although John et al (2007) did not measure the deeper abdominal muscles, it seems reasonable that they were not as active as *external oblique* during isometric trunk rotation. Finally, the fact that John et al (2007) felt that they could not pool data between participants also highlights the high degree of inter-participant variability in the relationship between changes in muscle thickness and muscle activity.

Timing of muscle activation: Most of the research linking dysfunction of *transversus abdominis* and *lumbar multifidus* to low back pain has focused on their delayed activation (Hodges and Richardson 1998, Hodges and Richardson 1999, Hodges 1999, Hodges 2001, Hungerford et al 2003). Until recently, however, timing of these deeper muscles could only be evaluated with invasive needle EMG. Two other ultrasound technologies (M-mode and tissue doppler imaging ultrasound) now show promise in the measurement of the timing muscle activation. All three studies investigating M-mode and tissue doppler imaging ultrasound found a relatively consistent delay of the ultrasound measures compared with EMG (Mannion et al 2008, Vasseljen et al 2006, Vasseljen et al 2009). This systematic delay is consistent with an expected electromechanical delay between the onset of electrical activity and the development of tension in the muscle (Mannion et al 2008). Although more research is needed in this area, the evidence so far suggests that ultrasound may be a useful measure of the timing of muscle activation as long as measurements are compared only to other ultrasound measurements (eg, between individuals or pre-post intervention) and/or corrected for the delay.

Construct validity

Construct validity is especially important in situations where no reference standard exists (eg, psychological constructs) (Streiner and Norman 2003). Because reference standards are readily available for the constructs that rehabilitative ultrasound imaging attempts to measure (MRI for morphology and EMG for function), we focused primarily on evaluating these criterion-related studies. Nonetheless, studies demonstrating construct validity of ultrasound strengthen the inferences that we can make from ultrasound measurements. For instance, the fact that several studies have demonstrated that patients with low back pain exhibit different trunk muscle morphology and function (both during voluntary and automatic tasks) than asymptomatic individuals, strengthens the argument that trunk muscle deficits are related to low back pain and should be addressed as part of rehabilitation.

Sensitivity to change

Although trunk muscles are thought to develop morphological and functional deficits after back injuries, and such deficits are assumed to be correctable with strengthening and motor control exercises, few studies were identified which quantify such changes. Only two studies, one on *transversus abdominis* (Akbari et al 2008) and one on *lumbar multifidus* (Hides et al 2008), quantified the longitudinal changes which occurred following stabilisation exercises with enough information for readers to be able to calculate a sensitivity to change index. Both studies investigated the size of the muscles at rest; therefore we have no information about measures of muscle activation and how sensitive they are to change. The only study which reported such information regarding decrements in muscle size or activation investigated immediate changes following a noxious intramuscular injection (Kiesel et al 2008). While such a study documents an important phenomenon, it is questionable to generalise the ability to detect short-term changes from artificially-induced pain to the ability to detect changes from naturally-occurring low back pain. Lastly, no studies were identified that compared changes in ultrasound measurements with external measurements of clinically-relevant outcomes (eg, pain or disability) or external measurements of importance (eg, global perceived effect). While minimum detectable changes can be derived from reliability studies, minimum clinically-important differences require responsiveness studies that include comparisons with external measurements indicating when

an important change has occurred (Portney & Watkins 2008).

Implications for practice

There is a growing body of evidence supporting the ability of rehabilitative ultrasound imaging to measure trunk muscle size and activation accurately during most isometric sub-maximal contractions. While clinicians can be confident of ultrasound measures of size (thickness and/or cross-sectional area), using change in size to reflect muscle activation requires more care. As expected from the complexity of the relationship between muscle size and muscle activation, the ability of ultrasound to measure muscle activation is probably context dependent. Clinicians can confidently use ultrasound to measure muscle activation during low levels of isometric contraction of *transversus abdominis*, *internal oblique*, *external oblique,* and *lumbar multifidus*. Clinicians should be careful when using ultrasound to measure muscle activation during high levels of contraction, during concentric or eccentric contractions, or during tasks that have not been validated. Finally, resting measures of trunk muscle size appear sensitive to both positive and negative change over time.

Limitations of the review

The task of selecting operational definitions to test the validity of rehabilitative ultrasound imaging and determining which studies fit these definitions is complex and somewhat subjective. Using slightly different operational definitions may have resulted in different inclusion and exclusion of studies. Although the studies selected for inclusion were evaluated by two independent reviewers, some conflicts had to be resolved by a third reviewer. Assessing the quality of the included studies was also difficult since there is no standard list of criteria to review methodological studies evaluating validity or sensitivity to change. Furthermore, none of the studies selected as demonstrating construct validity and sensitivity to change were primarily aimed at 'establishing validity'. Since the primary purpose and design of these studies varied greatly, the tool used to assess their quality had to be very general.

Recommendations for future research

Since the validity of using rehabilitative ultrasound imaging to quantify muscle activation seems to be context-dependent and differs depending on the muscle involved, the level of muscle contraction, and the contraction strategy utilized, more validation research is needed. Studies comparing changes in *lumbar multifidus* muscle size with EMG activity need to investigate different contraction strategies in wider ranges of muscle contractions. Studies comparing changes in abdominal muscle size with EMG activity need to be replicated to determine the relationship between ultrasound and EMG measures, as well as to investigate different contraction strategies. Finally, the sensitivity to change and responsiveness of ultrasound measures requires further elucidation. Studies are needed to investigate the rate of deterioration and extent of atrophy and activation deficits following the development of low back pain. More longitudinal studies documenting changes in muscle size and activation alongside clinical outcomes resulting from training are needed in order to determine what constitutes clinically-important change. ■

eAddenda: Appendix 1 and Appendix 2 available at AJP. www.physiotherapy.asn.au

Disclaimer: Shane Koppenhaver is a PhD Candidate at the University of Utah and a Major in the US Army. The opinions or assertions contained herein are the private views of the author and are not to be construed as official or as reflecting the views of the Departments of the Army or Defense.

Correspondence: Shane Koppenhaver, Physical Therapy Department, University of Utah, Salt Lake City, Utah, USA. Email: shanekoppenhaver@mac.com

References

Akbari A, Khorashadizadeh S, Abdi G (2008) The effect of motor control exercise versus general exercise on lumbar local stabilizing muscles thickness: randomized controlled trial of patients with chronic low back pain. *Journal of Back and Musculoskeletal Rehabilitation* 21: 105–112.

Bemben MG (2002) Use of diagnostic ultrasound for assessing muscle size. *Journal of Strength and Conditioning Research* 16: 103–108.

Boyett M R, Frampton J E, Kirby M S (1991) The length, width and volume of isolated rat and ferret ventricular myocytes during twitch contractions and changes in osmotic strength. *Experimental Physiology*, *76:* 259–270.

Coldron Y, Stokes M, Cook K (2003) Lumbar multifidus muscle size does not differ whether ultrasound imaging is performed in prone or side lying. *Manual Therapy* 8: 161–165.

Cook C, Hegedus E (2007) Orthopedic Physical Examination Tests: An Evidence-based Approach. Upper Saddle River: Pearson Prentice Hall.

Critchley D (2002) Instructing pelvic floor contraction facilitates transversus abdominis thickness increase during low-abdominal hollowing. *Physiotherapy Research International* 7: 65–75.

Critchley DJ, Coutts FJ (2002) Abdominal muscle function in chronic low back pain patients: measurement with real-time ultrasound scanning. *Physiotherapy* 88: 322–332.

Endleman I, Critchley DJ (2008) Transversus abdominis and obliquus internus activity during pilates exercises: measurement with ultrasound scanning. *Archives of Physical Medicine and Rehabilitation* 89: 2205–2212.

Ferreira PH, Ferreira ML, Hodges PW (2004) Changes in recruitment of the abdominal muscles in people with low back pain: ultrasound measurement of muscle activity. *Spine* 29: 2560–2566.

Hebert J, Koppenhaver S, Parent E, Fritz J (2009) A systematic review of the reliability of rehabilitative ultrasound imaging for the quantitative assessment of the abdominal and lumbar trunk muscles. *Spine*: In press.

Herbert RD, Moseley AM, Butler JE, Gandevia SC (2002) Change in length of relaxed muscle fascicles and tendons with knee and ankle movement in humans. *Journal of Physiology* 539: 637–645.

Hides J, Wilson S, Stanton W, McMahon S, Keto H, McMahon K, Bryant M, Richardson C (2006) An MRI investigation into the function of the transversus abdominis muscle during 'drawing-in' of the abdominal wall. *Spine* 31: E175–178.

Hides JA, Jull GA, Richardson CA (2001) Long-term effects of specific stabilizing exercises for first-episode low back pain. *Spine* 26: E243–248.

Hides JA, Richardson CA, Jull GA (1995) Magnetic resonance imaging and ultrasonography of the lumbar multifidus muscle: comparison of two different modalities. *Spine* 20: 54–58.

Hides JA, Stokes MJ, Saide M, Jull GA, Cooper DH (1994) Evidence of lumbar multifidus muscle wasting ipsilateral to

symptoms in patients with acute/subacute low back pain. *Spine* 19: 165–172.

Hides JA, Wong I, Wilson SJ, Belavy DL, Richardson CA (2007) Assessment of abdominal muscle function during a simulated unilateral weight–bearing task using ultrasound imaging. *Journal of Orthopedic and Sports Physical Therapy* 37: 467–471.

Hides JA, Belavy DL, Cassar L, Williams M, Wilson S J, Richardson CA (2009) Altered response of the anterolateral abdominal muscles to simulated weight-bearing in subjects with low back pain. *European Spine Journal* 18: 410–418.

Hides JA, Richardson CA, Jull GA (1996) Multifidus muscle recovery is not automatic after resolution of acute, first-episode low back pain. *Spine* 21: 2763–2769.

Hides JA, Stanton WR, McMahon S, Sims K, Richardson CA (2008) Effect of stabilization training on multifidus muscle cross-sectional area among young elite cricketers with low back pain. *Journal of Orthopedic and Sports Physical Therapy* 38: 101–108.

Hodges P, Holm AK, Hansson T, Holm S (2006) Rapid atrophy of the lumbar multifidus follows experimental disc or nerve root injury. *Spine* 31: 2926–2933.

Hodges PW (1999) Is there a role for transversus abdominis in lumbo-pelvic stability? *Manual Therapy* 4: 74–86.

Hodges PW (2001) Changes in motor planning of feedforward postural responses of the trunk muscles in low back pain. *Experimental Brain Research*, *141:* 261–266.

Hodges PW (2005) Ultrasound imaging in rehabilitation: just a fad? *Journal of Orthopedic and Sports Physical Therapy* 35: 333–337.

Hodges PW, Pengel LH, Herbert RD, Gandevia SC (2003) Measurement of muscle contraction with ultrasound imaging. *Muscle and Nerve* 27: 682–692.

Hodges PW, Richardson CA (1998) Delayed postural contraction of transversus abdominis in low back pain associated with movement of the lower limb. *Journal of Spinal Disorders* 11: 46–56.

Hodges PW, Richardson CA (1999) Altered trunk muscle recruitment in people with low back pain with upper limb movement at different speeds. *Archives of Physical Medicine and Rehabilitation* 80: 1005–1012.

Hollingworth W, Medina LS, Lenkinski RE, Shibata DK, Bernal B, Zurakowski D, Comstock B, et al (2006) Interrater reliability in assessing quality of diagnostic accuracy studies using the QUADAS tool. A preliminary assessment. *Academic Radiology* 13: 803–810.

Hungerford B, Gilleard W, Hodges P (2003) Evidence of altered lumbopelvic muscle recruitment in the presence of sacroiliac joint pain. *Spine* 28: 1593–1600.

Ishida Y, Kanehisa H, Kondo M, Fukunaga T, Carroll JF, Pollock ML et al (1994) Body fat and muscle thickness in Japanese and Caucasian females. *American Journal of Human Biology* 6: 711–718.

Ito M, Kawakami Y, Ichinose Y, Fukashiro S, Fukunaga T (1998) Nonisometric behavior of fascicles during isometric contractions of a human muscle. *Journal of Applied Physiology* 85: 1230–1235.

John EK, Beith ID (2007) Can activity within the external abdominal oblique be measured using real–time ultrasound imaging? *Clinical Biomechanics* 22: 972–979.

Kiesel KB, Uhl T, Underwood FB, Nitz AJ (2008) Rehabilitative ultrasound measurement of select trunk muscle activation during induced pain. *Manual Therapy* 13: 132–138.

Kiesel KB, Uhl TL, Underwood FB, Rodd DW, Nitz AJ (2007a) Measurement of lumbar multifidus muscle contraction with rehabilitative ultrasound imaging. *Manual Therapy* 12: 161–166.

Kiesel KB, Underwood FB, Matacolla C, Nitz AJ, Malone TR (2007b) A comparison of select trunk muscle thickness change between subjects with low back pain classified in the treatment–based classification system and asymptomatic controls. *Journal of Orthopedic and Sports Physical Therapy* 37: 596–607.

Lee SW, Chan CK, Lam TS, Lam C, Lau NC, Lau RW, Chan ST (2006) Relationship between low back pain and lumbar multifidus size at different postures. *Spine* 31: 2258–2262.

Mannion AF, Pulkovski N, Schenk P, Hodges PW, Gerber H, Loupas T et al(2008) A new method for the noninvasive determination of abdominal muscle feedforward activity based on tissue velocity information from tissue Doppler imaging. *Journal of Applied Physiology* 104: 1192–1201.

Masuda T, Miyamoto K, Oguri K, Matsuoka T, Shimizu K (2005) Relationship between the thickness and hemodynamics of the erector spinae muscles in various lumbar curvatures. *Clinical Biomechanics* 20: 247–253.

McMeeken JM, Beith ID, Newham DJ, Milligan P, Critchley DJ (2004) The relationship between EMG and change in thickness of transversus abdominis. *Clinical Biomechanics* 19: 337–342.

Misuri G, Colagrande S, Gorini M, Iandelli I, Mancini M, Duranti R et al (1997) In vivo ultrasound assessment of respiratory function of abdominal muscles in normal subjects. *European Respiratory Journal* 10: 2861–2867.

Narici MV, Binzoni T, Hiltbrand E, Fasel J, Terrier F, Cerretelli P (1996) In vivo human gastrocnemius architecture with changing joint angle at rest and during graded isometric contraction. *Journal of Physiology 496:* 287–297.

Oguri K, Zhao L, Du N, Kato Y, Miyamoto K, Hayakawa M (2004) Association of habitual long-distance running with the thickness of skeletal muscles and subcutaneous fat in the body extremities and trunk in middle-aged men. *Journal of Sports Medicine and Physical Fitness* 44: 517–423.

Otten E (1988) Concepts and models of functional architecture in skeletal muscle. *Exercise and Sport Science Review* 16: 89–137.

Patrick DL, Chiang YP (2000) Measurement of health outcomes in treatment effectiveness evaluations: conceptual and methodological challenges. *Medical Care* 38: II 14–25.

Portney LG, Watkins MP (2008) Foundations of Clinical Research: Applications to Practice (3rd Edition). Upper Saddle River: Pearson Prentice Hall.

Price TB, Kamen G, Damon BM, Knight CA, Applegate B, Gore JC et al (2003) Comparison of MRI with EMG to study muscle activity associated with dynamic plantar flexion. *Journal of Magnetic Resonance Imaging* 21: 853–861.

Rankin G, Stokes M, Newham DJ (2006) Abdominal muscle size and symmetry in normal subjects. *Muscle and Nerve* 34: 320–326.

Reeve A, Dilley A (2009) Effects of posture on the thickness of transversus abdominis in pain-free subjects. *Manual Therapy*. doi:10.1016/j.math.2009.02.008

Shrout PE (1998) Measurement reliability and agreement in psychiatry. *Statistical Methods in Medical Research* 7: 301–317.

Smidt N, Deeks J, Moore TE (2008) Cochrane Handbook for Diagnostic Test Accurcy Reviews [Chapter 4]. The Cochrane Collaboration, 2008. (Available from http://srdta.cochrane.org/en/authors.html).

Springer BA, Mielcarek BJ, Nesfield TK, Teyhen DS (2006) Relationships among lateral abdominal muscles, gender, body mass index, and hand dominance. *Journal of Orthopaedic and Sports Physical Therapy 36:* 289–297.

Stanton T, Kawchuk G (2008) The effect of abdominal stabilization contractions on posteroanterior spinal stiffness. *Spine* 33: 694–701.

Stokes M, Rankin G, Newham DJ (2005) Ultrasound imaging of lumbar multifidus muscle: normal reference ranges for measurements and practical guidance on the technique.

Manual Therapy 10: 116–126.

Stratford PW, Binkley FM, Riddle DL (1996) Health status measures: strategies and analytic methods for assessing change scores. *Physical Therapy* 76: 1109–1123.

Streiner DL, Norman GR (2003). *Health Measurement Scales: A Practical Guide to Their Development and Use.* New York: Oxford University Press.

Takemori S (1990) Influence of osmotic swelling on cross section and resting tension in isolated skeletal muscle fibers. *Japanese Journal of Physiology* 40: 595–611.

Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J et al (2007) Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology* 60: 34–42.

Teyhen DS (2006) Rehabilitative ultrasound imaging symposium San Antonio, TX, May 8–10, 2006. *Journal of Orthopaedic and Sports Physical Therapy* 36: A1–3.

Teyhen DS (2007) Rehabilitative ultrasound imaging: the roadmap ahead. *Journal of Orthopaedic and Sports Physical Therapy* 37: 431–433.

Van K, Hides JA, Richardson CA (2006) The use of real-time ultrasound imaging for biofeedback of lumbar multifidus muscle contraction in healthy subjects. *Journal of Orthopaedic and Sports Physical Therapy* 36: 920–925.

Vasseljen O, Fladmark AM, Westad C, Torp HG (2009) Onset in abdominal muscles recorded simultaneously by ultrasound imaging and intramuscular electromyography. *Journal of Electromyography and Kinesiology* 19: e23–31.

Vasseljen O, Dahl HH, Mork PJ, Torp HG (2006) Muscle activity onset in the lumbar multifidus muscle recorded simultaneously by ultrasound imaging and intramuscular electromyography. *Clinical Biomechanics* 21: 905–913.

Wallwork TL, Stanton WR, Freke M, Hides JA (2008) The effect of chronic low back pain on size and contraction of the lumbar multifidus muscle. *Manual Therapy* doi:10.1016/j.math.2008.09.006

Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J (2003) The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Medical Research Methodology* 3: 25.

Whiting PF, Weswood ME, Rutjes AW, Reitsma JB, Bossuyt PN, Kleijnen J (2006) Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. *BMC Medical Research Methodology* 6: 9.

Whittaker JL (2008) Ultrasound imaging of the lateral abdominal wall muscles in individuals with lumbopelvic pain and signs of concurrent hypocapnia. *Manual Therapy* 13: 404–410.

Yoshihara K, Shirai Y, Nakayama Y, Uesaka S (2001) Histochemical changes in the multifidus muscle in patients with lumbar intervertebral disc herniation. *Spine* 26: 622–626.

Zhao WP, Kawaguchi Y, Matsui H, Kanamori M, Kimura T (2000) Histochemistry and morphology of the multifidus muscle in lumbar disc herniation: comparative study between diseased and normal sides. *Spine* 25: 2191–2199.