ORIGINAL ARTICLE

Muscle thickness changes during abdominal hollowing: an assessment of between-day measurement error in controls and patients with chronic low back pain

**Anne F. Mannion · Natascha Pulkovski · Deborah Gubler ·
Mark Gorelick · David O'Riordan · Thanasis Loupas ·
Peter Schenk · Hans Gerber · Haiko Sprott**

Abstract Spine stabilization exercises, in which patients are taught to perform isolated contractions of the transverses abdominus (TrA) during "abdominal hollowing", are a popular physiotherapeutic treatment for low back pain (LBP). Successful performance is typically judged by the relative increase in TrA thickness compared with that of the internal (OI) and external (OE) oblique muscles, measured using ultrasound. The day-to-day measurement error (imprecision) associated with these indices of preferential activation has not been assessed but is important to know since it influences the interpretation of changes after treatment. On 2 separate days, 14 controls and 14 patients with chronic LBP (cLBP) performed abdominal hollowing exercises in hook-lying, while M-mode ultrasound images superimposed with tissue Doppler imaging (TDI) data were recorded from the abdominal muscles ($N = 5$ on each side). The fascial lines bordering the TrA, OI and OE were digitized, and muscle thicknesses were calculated. The between-day error (intra-observer) was expressed as the standard error of measurement, SEM; SEM as a percentage of the mean gave the coefficient of variation (CV). There were no significant between-day differences for the mean values of resting or maximal thickness for any muscle, in either group ($P > 0.05$). The median SEM and CV of all thickness variables was 0.71 mm and 10.9%, respectively for the controls and 0.80 mm or 11.3%, respectively for the cLBP patients. For the contraction ratios (muscle thickness contracted/ thickness at rest), the CVs were 3–11% (controls) and 5–12% (patients). The CVs were unacceptably high (30–50%, both groups) for the TrA preferential activation ratio (TrA proportion of the total lateral abdominal muscle thickness when contracted minus at rest). In both the controls and patients, the precision of measurement of absolute muscle thickness and relative change in thickness during abdominal hollowing was acceptable, and commensurate with that typical of biological measurements. The TrA preferential activation ratio is too imprecise to be of clinical use. Knowledge of the SEM for these indices is essential for interpreting the clinical relevance of any changes observed following physiotherapy.

Keywords Abdominal muscles · Physiotherapy exercises · Back pain · Reliability · Measurement error

A. F. Mannion (✉) · M. Gorelick · D. O'Riordan
Spine Center Division, Department of Research
and Development, Schulthess Klinik, Lengghalde 2,
8008 Zurich, Switzerland
e-mail: anne.mannion@kws.ch

N. Pulkovski · D. Gubler · P. Schenk · H. Sprott
Department of Rheumatology and Institute of Physical
Medicine, University Hospital Zürich, Zurich, Switzerland

T. Loupas
Philips Medical Systems, Athens, Greece

H. Gerber
Department of Mechanical and Process Engineering,
Institute for Biomechanics, Swiss Federal Institute
of Technology, Zurich, Switzerland

Introduction

In recent years, spine stabilization exercises have become an increasingly popular treatment for low back pain (LBP) [9, 22, 25] with evidence of their efficacy being provided

by two systematic reviews [9, 22, 25]. The first stage in the therapy process typically involves teaching the patient to perform sustained, isolated contractions of the deep-lying abdominal muscle, transversus abdominis (TrA), using "abdominal hollowing" (AH) exercises. Success in performing these exercises is given by the ability to selectively activate TrA in preference to the more superficial abdominal muscles, obliquus internus (OI) and obliquus externus (OE) and/or rectus abdominus [2, 25, 29]. Preferential TrA activation was originally assessed by examining the accompanying pressure changes recorded by an air-filled pressure bag attached to a sphygmomanometer gauge positioned under the abdomen of the prone patient [13, 25]. However, the validity of the pressure-sensor device for the quantification of TrA contraction has subsequently been challenged, since other movement strategies appear to elicit the same pressure changes [5] and the between-day reliability of measures obtained with this device is poor [28].

Ultrasound imaging has now superseded the pressure-sensor device for the assessment of deep trunk muscle activation and is being used with increasing frequency in both the research and clinical environment [12, 14]. Typically, assessment involves examination of the relative change in TrA muscle thickness compared with that of OI and OE [29].

A recent review of the use of ultrasound for assessing deep abdominal muscle activation in connection with LBP emphasized that, before being implemented in clinical practice, the variables measured must be shown to display adequate clinimetric properties [14]. Previous studies have shown that changes of muscle thickness, measured with ultrasound, are a valid index of muscle contraction as compared with electromyography [15, 20] or MRI indices of muscle thickness and the cross-sectional area of the abdomen [11]. Further, a number of studies have documented good test–retest reliability for static measures of individual abdominal muscle thicknesses at rest or in various contracted states [4, 6, 15, 21]. This is encouraging, since it shows that the measurement method itself is reliable under given, stable conditions; however, it is of limited relevance when determining whether the methodology is of use in clinical practice, where the key issue concerns the reliable measurement of the ability to contract the TrA muscle i.e., reliability of measures of muscle thickness change. The extent of the measurement error associated with the latter will effectively govern whether the method can be used to reliably classify someone with "trunk muscle dysfunction", and to detect real improvements/deterioration in function over time. Despite its frequent use in the clinical setting, no studies have previously examined the day-to-day measurement error associated with indices of TrA preferential activation; indeed, until recently [29], no clear criteria

even existed for assessing whether selective activation had actually been achieved.

Teyhen et al. [29] recently proposed various potentially useful indices for describing the activation of the trunk muscles during abdominal hollowing: the thickness ratio of TrA contracted to TrA rest; the thickness ratio of OI + OE (combined) contracted to OI + OE rest; and the difference in the TrA proportion of the total lateral abdominal muscle thickness (TrA/TrA + OI + OE) in going from the resting to the contracted state. The aim of the present study was to examine the intra-observer, between-day reliability of these indices of abdominal muscle activation, in a group of healthy LBP-free controls and in patients with chronic non-specific LBP (cLBP).

Methods

Study participants

A total of 14 healthy controls (7 male, 7 female) and 14 patients with cLBP (7 male, 7 female) took part. The mean ($\pm$SD) age, height, body mass and body mass index was 31 ± 10 years, 1.77 ± 0.10 m, 68 ± 14 kg, 21.8 ± 2.6 kg m^{-2}, respectively for the controls and 46 ± 9 years, 1.70 ± 0.05 m, 72 ± 6 kg, 24.9 ± 2.4 kg m^{-2}, respectively for the cLBP patients. The controls were recruited from the local universities/hospitals and had to have been LBP-free for the last year and have no history of LBP requiring medical attention or absence from work in the last 10 years. The patients were recruited from the local University hospitals. The inclusion criteria were: persistent LBP with or without referred pain (of a non-radical nature) for at least 3 months, serious enough to require medical attention or absence from work; average pain intensity over the last 2 weeks ≥ 3 and ≤ 8 on a 0–10 visual analogue scale; and willingness to comply with the study protocol. Exclusion criteria included: persistent severe pain, non-mechanical LBP, neurological symptoms, severe spinal instability, osteoporosis, structural deformity, systemic inflammatory disease, a decompensated metabolic disease or any other corresponding disorders preventing active rehabilitation, previous spinal fusion, severe cardiovascular diseases, acute infection, recent abdominal surgery, uncontrolled alcohol/drug abuse, and decompensated psychopathological diseases. Further exclusion criteria for both groups included pregnancy within the last 2 years. The study was approved by the local medical ethics committee and was a sub-study of a registered clinical trial (ISRCTN85021654). All participants gave their signed informed consent to participate after receiving verbal and written information about the study.

Test protocol

The volunteers visited the laboratory on two occasions, 1–2 weeks apart (mean 7 ± 2 days) and were assessed by the same investigator each time. Abdominal hollowing exercises were performed in the supine hook-lying position (hips in 30° flexion), by slowly contracting the abdominals to draw in the abdomen, and holding for 5 s. Subjects were instructed to first breathe in and then build up the contraction during expiration, breathing normally during the 5 s hold. They received a practice session (5–15 min), using ultrasound as a biofeedback tool [10, 12].

Ten repeated hollowing exercises were then performed (5 times with the transducer over the right abdominal muscles, and 5 times over the left abdominals) with a 1–2 min rest period between each. No verbal feedback or biofeedback was given during the actual test. Subjects were asked not to practice before their next test a week later; just before the latter, they were reminded how to perform the exercise and were allowed a brief practice (without biofeedback).

Ultrasound recordings

Ultrasound images were recorded at 333 Hz using a Philips HDI-5000 (Philips Medical Systems, Bothell, WA, USA) with a linear-array transducer (5–12 MHz); the images were superimposed with tissue Doppler image (TDI) data.

Using B (brightness)-mode ultrasound, the transducer was positioned 2.5 cm anteromedial to the mid-point between the iliac crest and the costal margin on the mid-axillary line, where the fascial boundaries between TrA, OI and OE and the superior edge of the TrA fascia lie parallel [21]. A 130 × 120 × 10 mm gel standoff pad (Sonar-Aid, Alloga AG, Burgdorf, Switzerland) and transmission gel

were placed between the transducer head and the skin. To ensure constant pressure and minimize relative movement between the transducer and abdomen during the tests, the transducer was housed in a high-density foam block, which was secured with Velcro straps around the pelvis.

M-mode recordings were made approximately 2–3 s prior to and throughout the 5 s abdominal hollowing manoeuvre. A printout of the ultrasound image was made at the start, and the muscle layers were labelled, to assist with the later orientation and identification of the muscle fascial borders during digitization.

The grey scale and TDI tissue velocity data from the M-mode ultrasound files, and event-marker data fed into the ultrasound machine's ECG channel (to indicate when the instruction was given to begin contraction) were exported in digital form using the ResearchLink option of the HDI-5000 system, and stored on computer.

Data processing

The leading edge points (i.e., the upper border) of the fascia of the muscle of interest were marked as manually selected control points at regular intervals throughout the M-mode image (black vertical bars in Fig. 1) and a custom-written plug-in of the HDI-Lab software (version 1.9 ATL/Philips Medical Systems, Bothell, WA, USA) was then used to automatically track the borders between adjacent control points, relying on the TDI velocity information to derive the displacement of a given point between two adjacent M-mode columns (displacement being equal to tissue velocity multiplied by the time difference between adjacent M-mode columns).

The distance between the top and bottom fascial lines for each M-mode column gave a measure of the thickness of the muscle over time. This was exported, as text data,
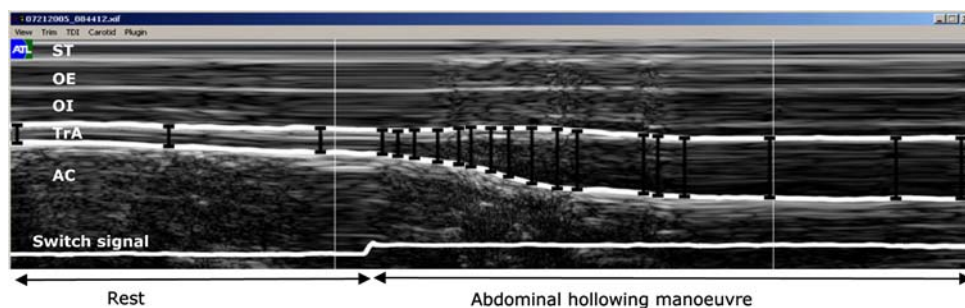

Fig. 1 M-mode ultrasound image of the abdominal hollowing manoeuvre. The distances between fascial borders were derived by means of a semi-automatic approach, based on manually selected control points (*vertical black bars*) plus tissue Doppler velocity information to track the borders between adjacent control points (shown here for TrA, transversus abdominis, as *thick white lines* bordering the muscle). *Note* for clarity, all markings are shown with

thicker line-widths than those used for the actual analysis process. No time or depth scales were displayed on the M-mode image during digitization; however, the image represents approximately 4 s worth of data (*x*-axis) (∼1.5 s of rest and ∼2.5 s of abdominal hollowing) with a total scan depth of ∼37 mm (*y*-axis). Switch signal: notch shows where instruction to begin was given. *ST*, subcutaneous tissue; *OE* obliquus externus; *OI* obliquus internus; *AC* abdominal contents

into a custom-written LabView software programme to determine: (1) the resting thicknesses of TrA, OI and OE (1 s value during quiet rest, just before the contraction); (2) the maximal thickness of TrA over any given 3 s period during the contraction; and (3) the thicknesses of OI and OE at the point of maximum TrA thickness (Fig. 2).

From the above data, the following indices were determined [29]:

(1) TrA contraction ratio = TrA thickness contracted/ TrA thickness at rest,
(2) OE + OI contraction ratio = OE + OI thickness contracted/OE + OI thickness at rest,
(3) TrA preferential activation ratio (difference in the TrA proportion of the total lateral abdominal muscle thickness in going from the relaxed to the contracted state) = (TrA contracted/TrA + OE + OI contracted) − (TrA at rest/TrA + OE + OI at rest).

The utility of a further index relating just to OE thickening was also investigated, since (in practice) co-activation of OI, but not OE, is sometimes considered acceptable during hollowing:

(1) OE contraction ratio = OE thickness contracted/OE thickness at rest.

To examine the error of measurement associated with the digitizing procedure itself, 330 images were measured by each of 2 differe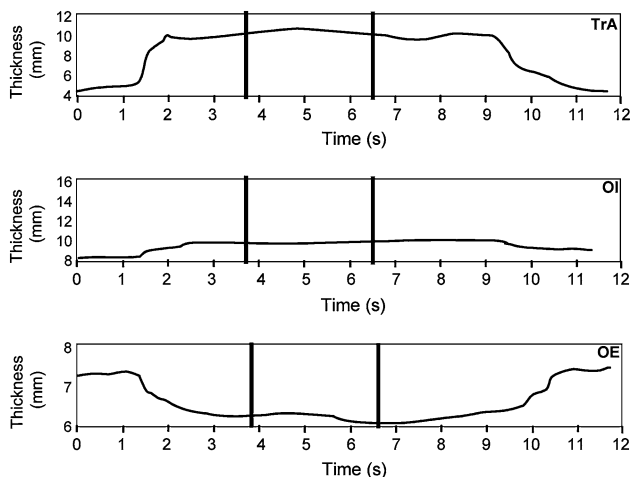nt people (selected at random from a group of 6 investigators who were all involved in the digitization of the main study data).

Data analysis/statistics

Right and left sides were considered as separate cases (data-sets), rather than taking an average of both, since previous studies have shown that abdominal muscle function can differ between body sides [11, 24]. Further, the data for the controls and the patients were examined separately, to assess whether the measurement error was comparable for the two groups. The mean values of the five trials for a given person on a given side and a given day were used to assess between-day reliability (because the average of multiple trials during a given test session improves the precision of measurement [16]).

For the assessment of between-day reliability (which comprised all sources of error: the biological error associated with an individual's repeat performance, the error of repositioning of the ultrasound transducer and the measurement error of digitization), the following statistics were determined from the output of a repeated measures analysis of variance: mean (SD) values; the significance of the difference between mean values; the intra-class correlation coefficient [ICC (3,1), i.e., two-way, mixed-effects [26] with results reported for single day reliability of the mean of five trials]; the standard error of measurement (SEM, given by the square root of the within subjects residual mean squares error and also referred as the "within subjects standard deviation" or "typical error of measurement" [16]) and the coefficient of variation (CV in %: SEM/mean value for the given parameter × 100). The SEM was reported in preference to the "limits of agreement", for the reasons detailed in Hopkins [16] and since the latter is so closely related to the SEM (being approximately equivalent to $3\times$ SEM).

The SEM is used to determine the degree of change required in a given individual's measure to establish that change (with a given level of confidence) as "real", over and above measurement error. This is sometimes referred to as "minimum (or smallest) detectable change", or MDC [3] and is especially useful for interpreting the relevance of any changes recorded after an intervention. At the 95% confidence level, the $MDC_{95\%}$ is calculated as $1.96 \times \sqrt{2} \times SEM$ (=2.77 × SEM) [3].

Similar reliability statistics (but using an ICC (1,1) [26]) were also calculated for the inter-examiner measurement error associated with the digitization procedure itself.

Statview 5.0 (SAS Institute, Cary, NC, USA) and SPSS v11.0 for Mac OS X (SPSS Inc, Chicago, IL, USA) were used for the statistical analyses. Significance was accepted at the 5% level.

Fig. 2 Abdominal muscle thicknesses, determined with a customized software programme, written in LabView. *TrA*, transversus abdominis; *OI*, obliquus internus; *OE* obliquus externus. The traces show the increasing thickness of TrA (trace 1) and OI (trace 2) during hollowing and the concomitant reduction in thickness of OE (trace 3). The resting muscle thickness is determined from the first 1 s period before contraction. The maximum value for TrA and the corresponding thicknesses of OI and OE at maximum TrA thickness are determined from the highest 3 s period over the contraction (indicated by *parallel bars*)

Results

Image quality

The tests yielded 560 ultrasound/TDI data files (28 subjects $\times$ 2 sides $\times$ 5 trials $\times$ 2 days). Sixty-two of these (11%) had to be disregarded due to poor image quality. For one (control) male on day one, this was the case for all files from the right abdominals and all but one file from the left abdominals; hence no day-to-day comparisons were possible for him (leaving $N = 13$ in the control group). For the remaining individuals, muscle thicknesses were calculated from a mean 4.7 (median 5.0) trials per body side.

Measurement error associated with muscle thickness digitization

The ICCs for the inter-observer reliability of measurement of muscle thickness (i.e., for re-digitization of a given set of files) were acceptably high, ranging from 0.79 (for OE at TrA_{max}) to 0.93 (for $TrA_{relaxed}$). The corresponding SEMs ranged from 0.29 mm (for $TrA_{relaxed}$) to 0.95 mm (for OI at TrA_{max}), and, when expressed as a percentage of the corresponding mean thicknesses, 7.1% (for TrA_{max}) to 12.2% (for $OI_{relaxed}$).

Muscle thickness at rest and during abdominal hollowing

There were no significant differences between test days for the mean values of absolute thickness for any muscle in either the control or the patient group (Tables 1, 2).

For the control group, the ICCs for the muscle thickness measures ranged from 0.26 (right $OE_{relaxed}$) to 0.94 (left $OI_{relaxed}$ and left OI at TrA_{max}), with a median value for all variables of 0.80 (Table 1). The ICCs for the muscle thickness measures in the cLBP patient group ranged from 0.31 (right OE at TrA_{max}) to 0.89 (right $TrA_{relaxed}$), with a median value for all variables of 0.77 (Table 2).

The standard error of measurement (SEM) for the muscle thickness measures in the control group ranged from 0.40 mm (left and right $TrA_{relaxed}$) to 1.03 mm (left $OE_{relaxed}$); respective values for the cLBP group were 0.27 mm (right TrA_{max}) to 1.25 mm (right OE at TrA_{max}). The CV for all muscle thickness variables was similar in both groups, ranging from 7 to 21% with a median value of 11% (Tables 1, 2).

Contraction ratios and TrA preferential activation index

The ICCs for the contraction ratios for TrA, OI + OE and OE ranged from 0.50 to 0.72 for the controls and from 0.25

to 0.80 for the cLBP patients (Tables 1, 2). The CVs were lower than those for the thickness measures, ranging from 3.2 to 11.4% for the controls and from 4.5 to 11.6% for the cLBP patients (for both groups, median value 5.6%).

The ICCs for the TrA preferential activation ratio were poor to moderate (0.32–0.62) and the CVs were high (30–50%; Tables 1, 2).

Discussion

General methodological considerations

The present study examined the intra-observer, between-day reliability of measures of deep trunk muscle thickness during abdominal hollowing in male and female volunteers who were either LBP-free or who suffered from non-specific [1] cLBP. We chose to examine the measurement error for both controls and patients, as the former was expected to reflect the best-case scenario—predominantly technical error and the normal day-to-day biological variation—whilst the patient group would better reflect the reality (giving external validity), should additional factors such as pain, motor control dysfunctions or fear of the test situation influence performance. It was not the purpose of the present study to compare the actual performance of the controls and the patients; this would have required larger groups of subjects, carefully matched in relation to various anthropometric variables that influence abdominal muscle thickness [24].

In contrast to previous studies [4, 6, 15, 21], the muscle thickness was recorded over the whole contraction period and both the resting level and the highest thickest of TrA during the contraction were determined automatically; this removed any "subjectivity" from the thickness analyses. Since the exercise test requires that the contraction be sustained, it was considered expedient to measure the maximum thickness over a 3 s period, thereby avoiding any transient peaks given by the instantaneous maximum.

In the assessment of human performance capacity, "one-off" measurements rarely provide sufficiently accurate data [16]. As such, in the present study, five trials were carried out per test setting (person, side, day) and these were averaged before further analysis of between-day reliability. One subject had to be eliminated from the analysis entirely, due to the consistently poor quality of his ultrasound images on one of the test days; the fact that the method cannot be applied successfully in all individuals, at all times, must be viewed as a limitation of the procedure.

The variability between trials performed on a given day is expected to incorporate both the natural variability in human performance plus the error of digitization of the fascial borders; for between-day measures, the positioning

Table 1 Reliability of measures of abdominal muscle thickness and contraction index scores on the two test days in the control group

	Side	Mean (day 1)	SD (day 1)	Mean (day 2)	SD (day 2)	ICC	P	SEM	CV (SEM as % mean)
TrA $_{rest}$(mm)	L	3.9	1.2	3.7	0.9	0.86	0.21	0.40	10.7
	R	4.0	1.0	3.7	0.9	0.83	0.12	0.40	10.3
TrA $_{max}$(mm)	L	5.5	1.4	5.4	1.2	0.75	0.66	0.65	12.0
	R	5.4	1.4	5.4	1.0	0.78	0.87	0.58	10.7
OI $_{rest}$(mm)	L	7.4	3.0	7.5	2.9	0.94	0.80	0.72	9.8
	R	6.4	2.1	6.7	2.1	0.92	0.15	0.58	8.8
OI at TrA $_{max}$(mm)	L	8.2	3.6	8.0	3.2	0.94	0.57	0.82	10.1
	R	6.9	2.6	7.2	2.5	0.91	0.36	0.78	11.0
OE $_{rest}$(mm)	L	5.3	1.5	5.2	1.7	0.59	0.84	1.03	19.6
	R	4.9	0.9	5.0	1.0	0.26	0.89	0.84	17.0
OE at TrA $_{max}$(mm)	L	5.0	1.2	5.1	1.5	0.58	0.82	0.88	17.3
	R	4.6	1.0	4.8	1.1	0.56	0.48	0.70	14.8
TrA contraction ratio	L	1.45	0.21	1.50	0.25	0.50	0.48	0.16	10.9
	R	1.39	0.26	1.48	0.21	0.52	0.15	0.16	11.4
OE + OI contraction ratio	L	1.05	0.05	1.04	0.06	0.61	0.69	0.03	3.2
	R	1.02	0.08	1.02	0.07	0.72	0.76	0.04	3.9
OE contraction ratio	L	0.97	0.08	1.00	0.10	0.60	0.21	0.06	5.8
	R	0.94	0.07	0.98	0.07	0.66	0.06	0.04	4.4
TrA preferential activation ratio	L	0.06	0.03	0.07	0.04	0.55	0.54	0.02	38.0
	R	0.06	0.03	0.07	0.03	0.62	0.12	0.02	30.2

TrA transversus abdominis; *OI* internal oblique; *OE* external oblique rest, resting thickness; *max* maximal thickness; *at TrA max* thickness at maximal TrA thickness; *L* left; *R* right *SD* standard deviation *ICC* intra-class correlation coefficient; *P* P value for the significance of the difference on day 1 and day 2; *SEM* standard error of measurement (within subjects standard deviation); *CV* coefficient of variation (SEM as % mean on the 2 days)

anew of the ultrasound transducer contributes additional measurement error. The between-day error is the more relevant index, since it assists in interpreting the clinical relevance of any changes observed after treatment; this was hence the focus of the present study.

Day-to-day reliability of muscle thickness measures

In the control group, the ICCs were acceptably high [7] for most of the thickness parameters (median 0.80); nonetheless, there were some measures, notably for OE, that showed only low to moderate ICCs. Interestingly, for OE (only), the validity of muscle thickness change as an indicator of muscle activity has been questioned in previous EMG validation studies [15, 18]. In the patient group, the ICCs were just slightly lower than those of the control group, with a median value of 0.77.

In practical terms, the standard error of measurement (SEM) ("typical error" [16]) delivers more useful information than the ICC [23]. In the present study, the SEMs for the various thickness measures were 0.40–1.03 mm for the controls and 0.27–1.25 mm for the cLBP patients; the

corresponding coefficients of variation (CV) were in the range 7–21% and were similar for the two groups. The CV is ideal for comparing the relative measurement error of variables with differing absolute values or units of measurement. For various biological and performance measurements, CVs of 10–20% are considered typical, with higher values being more common for performance variables than anatomical/biological variables [8, 17, 19, 27]. Our thickness values are hence commensurate with these, with a median CV of 11%. Interestingly, the CV in the present study for the between-day reliability was similar to that reported by Teyhen et al. [29] for the same-day inter-image reliability of TrA thickness and whole abdominal mass thickness (11–14%). This suggests that the greatest proportion of the day-to-day measurement error is associated with the measurement procedure itself.

A SEM of approximately 0.38 mm (average of right and left sides, patients and controls) for between-day measures of resting TrA thickness, and 0.70 mm for resting OI, would give minimum detectable change (MDC$_{95\%}$) values (see "Methods") of 1.1 and 1.9 mm, respectively. These are quite comparable to values previously reported in the literature for resting muscle, despite the slightly different

Table 2 Reliability of measures of abdominal muscle thickness and contraction index scores on the two test days in the cLBP group

Muscle and state	Side	Mean (day 1)	SD (day 1)	Mean (day 2)	SD (day 2)	ICC	P	SEM	CV (SEM as % mean)
TrA $_{rest}$(mm)	L	4.1	0.7	4.0	0.8	0.63	0.60	0.46	11.5
	R	3.9	0.8	3.7	0.8	0.89	0.07	0.27	7.2
TrA $_{max}$(mm)	L	5.5	0.9	5.4	1.1	0.41	0.83	0.78	14.3
	R	5.3	1.2	5.3	1.1	0.88	0.87	0.41	7.7
OI $_{rest}$(mm)	L	6.9	1.6	7.0	1.8	0.85	0.84	0.68	9.8
	R	7.6	1.7	7.6	1.4	0.73	0.99	0.82	10.8
OI at TrA $_{max}$(mm)	L	7.4	1.9	7.3	2.1	0.84	0.68	0.81	11.0
	R	8.1	1.7	8.1	1.6	0.74	0.99	0.82	10.1
OE $_{rest}$(mm)	L	5.6	1.3	5.9	1.1	0.51	0.43	0.84	14.6
	R	6.1	1.6	6.5	1.5	0.42	0.48	1.20	19.1
OE at TrA $_{max}$(mm)	L	5.6	1.3	5.6	1.0	0.59	0.93	0.76	13.6
	R	5.9	1.5	6.1	1.5	0.31	0.71	1.25	21.0
TrA contraction ratio	L	1.36	0.10	1.39	0.25	0.28	0.63	0.16	11.6
	R	1.37	0.19	1.47	0.19	0.80	0.01	0.09	6.0
OE + OI contraction ratio	L	1.03	0.06	1.00	0.07	0.39	0.10	0.05	5.6
	R	1.02	0.06	1.01	0.05	0.25	0.31	0.05	4.5
OE contraction ratio	L	0.99	0.08	0.95	0.09	0.57	0.10	0.05	5.6
	R	0.96	0.07	0.93	0.06	0.43	0.15	0.05	5.4
TrA preferential activation ratio	L	0.05	0.02	0.06	0.04	0.32	0.44	0.03	49.5
	R	0.06	0.02	0.07	0.02	0.48	0.07	0.02	27.4

TrA transversus abdominis; *OI* internal oblique; *OE* external oblique rest, resting thickness; *max* maximal thickness; *at TrA max* thickness at maximal TrA thickness; *L* left; *R* right; *SD* standard deviation; *ICC* intra-class correlation coefficient; *P* P value for the significance of the difference on day 1 and day 2; *SEM* standard error of measurement (within subjects standard deviation); *CV* coefficient of variation (SEM as % mean on the 2 days)

methodology used: Hodges et al. [15] reported MDC$_{95\%}$ values of 1.0 and 1.9 mm for TrA and OI, respectively, whilst Critchley et al. [6] reported somewhat lower values of 0.6 and 1.2 mm for resting TrA and OI, respectively. In examining absolute thickness measures in supine, standing and walking positions, Bunce et al. [4] reported MDC 95% values (for TA only) ranging from 0.9 to 1.8 mm.

Despite the fact that the relevance of the hollowing manoeuvre resides in the ability to preferentially thicken the deep-lying abdominals (in particular, the TrA) there are no reports in the literature regarding the SEMs associated with between-day measures of thickness during the abdominal hollowing test itself. In the present study, the SEMs for contracted muscle (approximately 0.61 and 0.81 mm for TrA and OI, respectively) yielded MDC$_{95\%}$ values of 1.7 and 2.2 mm, which are similar to those recorded at rest and hence appear to be acceptable from the perspective of human performance measurements [27]. Nonetheless, to determine whether this is an acceptable level of error in practice, clinical studies showing the typical increases in maximal thickness that are achievable/detectable through interventions such as spine stabilization exercises are still required.

Day-to-day reliability of contraction ratios and TrA preferential activation index

Potentially the most relevant indices of performance capacity during abdominal hollowing are those recently proposed by Teyhen et al. [29] to indicate the relative increase in thickness ("contraction ratio") of the deep-lying abdominal muscles and the preferential activation of TrA. In the present study, the ICCs for the contraction ratios were not particularly high. This may be because these indices effectively "normalize" the absolute thickness data by expressing the contracted thickness as a ratio of the resting thickness. Hence they remove a large proportion of the between-subjects' variance in absolute thickness, otherwise introduced by variations in body size [24]. The ICC is strongly dependent on the variance between subjects in the group under investigation [16, 23] and when the latter variance is low, then so too is the ICC. Indeed, the over-interpretation of ICCs in such circumstances has been cautioned against before [23] and it is recommended that the focus instead be placed on the size of the SEM, or the latter expressed as a proportion of the mean (i.e., the CV).

For all the indices apart from the TrA preferential activation ratio [29], the SEMs and the corresponding CVs were very low, making them promising measures for further investigations. Since they represent normalized values, they constitute a potentially valuable measure in studies where body size could otherwise play a confounding role. Further studies involving cross-sectional comparisons of large groups of individuals are recommended to examine whether the indices are able to differentiate between those with and without a history of LBP. Further, prospective studies should examine whether they are more sensitive to change than absolute measures of muscle thickness or thickness change. Finally, future studies should assess the relationship between improvements in function as measured by these indices and improvements in clinical symptoms (pain, disability) to confirm that the test procedure of abdominal hollowing is indeed relevant as a clinical tool for assessment, diagnosis and outcome measurement.

Of all the thickness measures and indices examined in the current study, the poor reliability (and in particular, the high CV) of the "TrA preferential activation ratio" renders it the least reliable measure, hence questioning its use in clinical practice. Perhaps a modification of the index or a combination-index based on the other contraction ratios would yield a more reliable measure of preferential activation during abdominal hollowing; this should be investigated in future studies.

References

1. Airaksinen O, Brox JI, Cedraschi C, Hildebrandt J, Klaber-Moffett J, Kovacs F, Mannion AF, Reis S, Staal JB, Ursin H, Zanoli G (2006) Chapter 4. European guidelines for the management of chronic nonspecific low back pain. Eur Spine J 15(Suppl 2):S192–S300

2. Allison GT, Kendle K, Roll S, Schupelius J, Scott Q, Panizza J (1998) The role of the diaphragm during abdominal hollowing exercises. Aust J Physiother 44:95–102

3. Beaton DE (2000) Understanding the relevance of measured change through studies of responsiveness. Spine 25:3192–3199

4. Bunce SM, Moore AP, Hough AD (2002) M-mode ultrasound: a reliable measure of transversus abdominis thickness? Clin Biomech (Bristol, Avon) 17:315–317

5. Cairns MC, Harrison K, Wright C (2000) Pressure biofeedback: a useful tool in the quantification of abdominal muscle dysfunction? Physiotherapy 86:127–138

6. Critchley DJ, Coutts FJ (2002) Abdominal muscle function in chronic low back pain patients. Physiotherapy 88:322–332

7. Currier DP (1990) Elements of research in physical therapy. Williams and Wilkins, Baltimore

8. Elfving B, Nemeth G, Arvidsson I, Lamontagne M (1999) Reliability of EMG spectral parameters in repeated measurements of back muscle fatigue. J Electromyogr Kinesiol 9:235–243

9. Ferreira PH, Ferreira ML, Maher CG, Herbert RD, Refshauge K (2006) Specific stabilisation exercise for spinal and pelvic pain: a systematic review. Aust J Physiother 52:79–88

10. Henry SM, Westervelt KC (2005) The use of real-time ultrasound feedback in teaching abdominal hollowing exercises to healthy subjects. J Orthop Sports Phys Ther 35:338–345

11. 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–E178

12. Hides JA, Jull GA, Richardson CA (1998) Use of real-time ultrasound imaging for feedback in rehabilitation. Man Ther 3:125–131

13. Hodges P, Richardson C, Jull G (1996) Evaluation of the relationship between laboratory and clinical tests of transversus abdominis function. Physiother Res Int 1:30–40

14. Hodges PW (2005) Ultrasound imaging in rehabilitation: just a fad? J Orthop Sports Phys Ther 35:333–337

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

16. Hopkins WG (2000) Measures of reliability in sports medicine and science. Sports Med 30:1–15

17. Howe T, Oldham J (1995) Functional tests in elderly osteoarthritic subjects: variability of performance. Nurs Stand 9:35–38

18. John EK, Beith ID (2007) Can activity within the external abdominal oblique be measured using real-time ultrasound imaging? Clin Biomech (Bristol, Avon) 22:972–979

19. Luoto S, Hupli M, Alaranta H, Hurri H (1996) Isokinetic performance capacity of trunk muscles. Part II: coefficient of variation in isokinetic measurement in maximal effort and in submaximal effort. Scand J Rehabil Med 28:207–210

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

21. Misuri G, Colagrande S, Gorini M, Iandelli I, Mancini M, Duranti R, Scano G (1997) In vivo ultrasound assessment of respiratory function of abdominal muscles in normal subjects. Eur Respir J 10:2861–2867

22. Rackwitz B, de Bie R, Limm H, von Garnier K, Ewert T, Stucki G (2006) Segmental stabilizing exercises and low back pain. What is the evidence? A systematic review of randomized controlled trials. Clin Rehabil 20:553–567

23. Rankin G, Stokes M (1998) Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. Clin Rehabil 12:187–199

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

25. Richardson C, Jull G, Hodges P, Hides J (1999) Therapeutic exercise for spinal stabilisation: scientific basis and practical techniques. Churchill Livingstone, Edinburgh

26. Shrout PE, Fleiss JL (1979) Intraclass correlations: uses in assessing rater reliability. Psychol Bull 86:420–428

27. Stokes M (1985) Reliability and repeatability of methods for measuring muscle in physiotherapy. Physiother Theory Prac 1:71–76

28. Storheim K, Bo K, Pederstad O, Jahnsen R (2002) Intra-tester reproducibility of pressure biofeedback in measurement of transversus abdominis function. Physiother Res Int 7:239–249

29. Teyhen DS, Miltenberger CE, Deiters HM, Del Toro YM, Pulliam JN, Childs JD, Boyles RE, Flynn TW (2005) The use of ultrasound imaging of the abdominal drawing-in maneuver in subjects with low back pain. J Orthop Sports Phys Ther 35:346–355