Warming up Ultrasound

20 GRAMS AT A TIME

The NEW Thermasonic® Gel Warmer For Single Use Packettes

Designed around 20 gram packettes of ultrasound gel, Parker's newest warmer adds comfort to our single use packettes' long list of benefits.

Like our other Thermasonic Gel Warmers, advanced microprocessor technology allows for custom temperature control, while a powerful heating element brings 20 packettes of gel to temperature quickly and efficiently. Engineered for performance, durability and control.

THERMASONIC® GEL WARMER

Warming Up Ultrasound

ISO 13485:2016

Learn more at: parkerlabs.com/thermasonic

PARKER | Parker Laboratories, Inc. | The sound choice in patient care.™ | 973.276.9500 | parkerlabs.com

AD-83-1 REV 0

Reliability of Real-time Ultrasound Imaging for the Assessment of Trunk Stabilizer Muscles

A Systematic Review of the Literature

Morteza Taghipour, PT, MSc, Mohammad Ali Mohseni-Bandpei, PT, PhD, Hamid Behtash, MD, Iraj Abdollahi, PT, PhD, Fatemeh Rajabzadeh, PT, MSc, Mohammad Reza Pourahmadi, PT, MSc, Mahnaz Emami, PT, MSc

Rehabilitative ultrasound (US) imaging is one of the popular methods for investigating muscle morphologic characteristics and dimensions in recent years. The reliability of this method has been investigated in different studies. As studies have been performed with different designs and quality, reported values of rehabilitative US have a wide range. The objective of this study was to systematically review the literature conducted on the reliability of rehabilitative US imaging for the assessment of deep abdominal and lumbar trunk muscle dimensions. The PubMed/MEDLINE, Scopus, Google Scholar, Science Direct, Embase, Physiotherapy Evidence, Ovid, and CINAHL databases were searched to identify original research articles conducted on the reliability of rehabilitative US imaging published from June 2007 to August 2017. The articles were qualitatively assessed; reliability data were extracted; and the methodological quality was evaluated by 2 independent reviewers. Of the 26 included studies, 16 were considered of high methodological quality. Except for 2 studies, all high-quality studies reported intraclass correlation coefficients (ICCs) for intra-rater reliability of 0.70 or greater. Also, ICCs reported for inter-rater reliability in high-quality studies were generally greater than 0.70. Among low-quality studies, reported ICCs ranged from 0.26 to 0.99 and 0.68 to 0.97 for intra- and inter-rater reliability, respectively. Also, the reported standard error of measurement and minimal detectable change for rehabilitative US were generally in an acceptable range. Generally, the results of the reviewed studies indicate that rehabilitative US imaging has good levels of both inter- and intra-rater reliability.

Key Words—abdominal muscles; low back pain; multifidus; musculoskeletal; reliability; systematic review; ultrasound

Low back pain (LBP) is one of the most common and costly health problems throughout the world.[1,2] About 40% of adult the population report LBP during their lifetime.[3–5] Trunk stabilizer muscles such as the deep abdominal and lumbar multifidus play a major role in providing stability and preventing LBP.[6] These muscles are considered key elements of lumbar stability.[6] Functional deficits and morphologic changes in trunk stabilizer muscles have been shown in patients with LBP.[7–10] Atrophy of paraspinal muscles is one of the common consequences of LBP.[11] Hides et al[12] reported that about 24 hours after onset of lumbar disk herniation, the cross-sectional area of the lumbar multifidus may decrease, and

Received November 7, 2017, from the Department of Physiotherapy (M.T., I.A., F.R., M.E.) and Pediatric Neurorehabilitation Research Center (M.A.M.-B.), University of Social Welfare and Rehabilitation Sciences, Tehran, Iran; University Institute of Physical Therapy, Faculty of Allied Health Sciences, University of Lahore, Lahore, Pakistan (M.A.M.-B.); Department of Orthopedics, Hazrat e Rasoul Akram Hospital, Iran University of Medical Sciences, Tehran, Iran (H.B.); and Department of Physiotherapy, School of Rehabilitation Sciences, Iran University of Medical Sciences, Tehran, Iran (M.R.P.). Manuscript accepted for publication March 20, 2018.

Address correspondence to Mohammad Ali Mohseni-Bandpei, PT, PhD, Pediatric Neurorehabilitation Research Center, University of Social Welfare and Rehabilitation Sciences, Koodakyar Street, Student Blvd., P.O. Box: 1985713834, Tehran, Iran.

E-mail: mohseni_bandpei@yahoo.com

Abbreviations

ICC, intraclass correlation coefficient; LBP, low back pain; US, ultrasound

doi:10.1002/jum.14661

consequently, paraspinal muscle atrophy may affect trunk proprioception.[13] Therefore, it seems that trunk muscle dimensions and function are important factors in the evaluation and treatment of patients with LBP.

There are a number of methods for evaluating muscle activity and morphologic characteristics, such as electromyography,[14] magnetic resonance imaging,[15] biopsy,[16] computed tomography,[17] and rehabilitative ultrasound (US) imaging.[18,19] Among these, rehabilitative US imaging is reported to be a low-risk, cost-effective, portable, and clinically accessible method compared to electromyography, computed tomography, biopsy, and magnetic resonance imaging.[20]

Reliability is one of the important psychometric properties of a measurement tool.[21] It is defined as "the degree to which measures are free from error and, therefore, yield consistent results."[22] Adequate reliability is necessary for accurate measurement using any tool or instrument.[21]

There are many studies that have evaluated the reliability of rehabilitative US imaging for measurement of trunk muscle dimensions. The reliability of rehabilitative US imaging for measurement of trunk muscles dimension has been reported to have a wide range, depending on the target population, patient's position, muscle state, and other influencing factors. Hebert et al[23] previously conducted a systematic review on the reliability of rehabilitative US imaging to evaluate abdominal and lumbar multifidus muscles from its beginning to May 2007. Therefore, this study was designed to systematically review published articles from June 2007 to August 2017 to determine whether rehabilitative US imaging is a reliable method for measuring abdominal and lumbar multifidus muscle dimensions in healthy individuals and patients with LBP. Furthermore, it also aimed to determine important parameters for rehabilitative US imaging to be clinically applied as a reliable, effective, and generalizable tool to measure and evaluate trunk muscle dimensions.

Methods

This systematic review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-analysis guidelines.[24]

Search Strategy

A search of the literature was performed in a systematic fashion, using the following key words: "reliability," "sonography," "ultrasound imaging," "trunk muscles," "multifidus," "transversus abdominis," "internal oblique," "within-day," "between-day," "inter-rater," and "intra-rater" from June 2007 to August 2017 in different electronic databases, including PubMed/MEDLINE, Scopus, Google Scholar, Science Direct, Embase, Physiotherapy Evidence, Ovid, and CINAHL.

After removing duplicate studies, 2 reviewers (M.T. and F.R.) screened titles and abstracts of all primary articles to determine relevant articles. Then, they evaluated the full text of potentially relevant articles independently to identify eligible articles based on inclusion/exclusion criteria (Table 1). If they did not reach an agreement, a third reviewer (M.A.M.-B.) assessed the eligibility of the articles. A Preferred Reporting Items for Systematic Reviews and Meta-analysis flowchart for the search strategy of this systematic review is presented in Figure 1.

Quality Assessment

A quality assessment of the included articles was performed by using the scale designed by Hebert et al.[23] This tool is a modified version of a quality assessment questionnaire developed by van Trijffel et al[25] for reviewing the methodological quality of reliability studies. It consists of 10 appraisal questions that evaluate the external validity (items 1–3), internal validity (items 4–8) and statistical method (items 9 and 10) of reliability studies (Table 2). All items were equally weighted and scored as "yes," "no," or "unclear" (yes = 1; no and unclear = 0). A higher score represents higher methodological quality. Furthermore, in the case of question 3, 5 parameters (including scanning point, amount of contact pressure, muscle state, transducer position, and participant's starting position) were used to determine the replicability of the assessment. Studies that met at least 4 of the 5 parameters obtained a score of yes for question 3.

Studies that received a score of yes on at least 50% of the items were considered high-quality studies.[25] Two reviewers (M.T. and F.R.) assessed the quality of all included studies independently. The level of agreement between the reviewers was calculated by another investigator using κ statistics (0–0.29, weak agreement; 0.30–0.59, moderate agreement; 0.60–0.89, good agreement; and 0.90–1, optimal agreement).[26] Furthermore, any disagreement between the reviewers was resolved by the third reviewer (M.A.M.-B).

Data Extraction

To perform descriptive analyses, data were extracted from the original included studies independently by the 2 reviewers (M.T. and F.R.). The extracted data contained the following information: authors, year, description of participants' characteristics (eg, sample size and demographics), rehabilitative US imaging (eg, model and transducer frequency), statistical analyses of reliability assessments (eg, intraclass correlation coefficient [ICC], standard error of measurement, and minimal detectable change). Then, the reviewers compared their results to achieve uniformity in the extracted data. If there were any disagreement between the reviewers, the final decision was made by the third reviewer (M.A.M.-B.).

Results

Search and Selection

A total of 250 published studies were identified through the electronic search (Figure 1). Of these, 67 duplicate articles were removed. Then 142 articles were excluded after title and abstract screening. Of the remaining 41 articles, 15 were excluded on the basis of full-text screening, and finally, 26 studies were considered eligible for inclusion in this review. Details of the included studies are provided in chronologic order from June 2007 to August 2017 in Table 3.

Quality Assessment/Data Extraction

The level of agreement between the reviewers seemed to be good ($\kappa = 0.73$). The final results of data extraction and a summary of reported ICCs in all included studies are presented in Tables 3 and 4, respectively. The results of the methodological quality assessment are presented in Table 4. As shown in Table 4, the scores of

Table 1. Selection Criteria

Inclusion	Exclusion
Published as a full article	If measurement properties were not assessed in the abdominal or lumbar region
Evaluate the reliability of B-mode ultrasound	Case studies, letters to the editor, and comments
Include imaging of the abdominal or lumbar trunk muscles	Published in a language other than English
Age of participants ≥ 18 y	

Table 2. Criteria for Assessing the Methodological Quality of Studies on Reliability

Criteria
1. Was a representative sample of participants used?
2. Was a representative sample of examiners used?
3. Is replication of the assessment procedure possible?
4. Were participants' characteristics under study stable during research?
5. Was an estimate of intraexaminer reliability sufficiently large?
6. Were examiners blinded to clinical information from participants?
7. Were examiners blinded to each other's results or to prior results?
8. Could nonrandom loss to follow-up be ruled out?
9. Were appropriate measures used for calculating reliability?
10. Were appropriate measures used for calculating precision?

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-analysis flowchart of included studies.

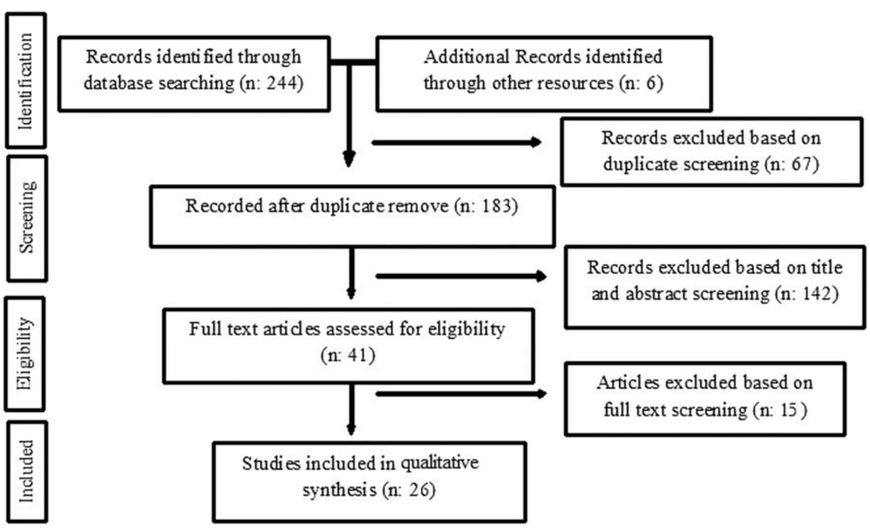

Table 3. Details of the Reviewed Studies in Chronologic Order From 2007 to 2017

Authors, y	Sample Size, Status	US Setting	Muscles, Dimensions, Status	Type of Reliability (Interval)	Reliability Estimate	Precision Estimate
Kiesel et al,[27] 2007[a]	15, LBP	SonoSite 180 Plus (SonoSite, Inc, Bothell, WA) computerized US unit with a 2–5-MHz curvilinear array transducer set at 5 MHz	TrA and MF thickness at rest and contraction	Intra-rater (same d)	$ICC_{3,3}$ TrA: rest, 0.98; contraction, 0.97 MF: rest, 0.99; contraction, 0.86–0.99	TrA, rest: SEM, 0.01; MDC, 0.02; contraction: SEM, 0.03; MDC, 0.06 MF, rest: SEM, 0.07; MDC, 0.19; contraction: SEM, 0.07–0.09; MDC, 0.19-0.25
Wallwork et al,[46] 2007	10, H	GE Diasonics Ultrasound, Inc (Santa Clara, CA) US unit with a 5-MHz curvilinear array transducer	MF thickness at rest	Inter-rater, intra-rater (same d)	Inter-rater $ICC_{2,1}$ Rest, 0.96–0.97 Intra-rater $ICC_{3,1}$ Rest, 0.88–0.95	MF: inter-rater SEM, 0.05–0.6; intra-rater SEM, 0.06–0.11
Teyhen et al,[43] 2008	10, H	Portable US unit (SonoSite 180 Plus) with a 60-mm, 2–5-MHz curvilinear array transducer	TrA and IO thickness at rest and contraction	Inter-rater (same d)	$ICC_{2,2} \geq 0.95$	TrA: SEM, 0.09 IO: SEM, 0.29
Costa et al,[28] 2009[a]	35, CLBP	Teratech (Terason Ultrasound Systems, Burlington, MA) US unit in B-mode with 10-cm, 5–10-MHz linear wide-band array transducer	TrA, IO, and EO thickness at rest and contraction	Intra-rater (2 d)	$ICC_{2,1}$ Thickness, 0.97	Thickness: SEM, 0.04 cm
Koppenhaver et al,[30] 2009[a]	30, NSLBP	SonoSite Titan US unit in B-mode with a 60-mm, 5-MHz curvilinear array transducer	TrA and MF thickness at rest and contraction	Intra-rater (same d)	$ICC_{3,k}$ TrA and MF: 0.79–0.96	TrA, contraction: SEM, 7.8%–19.2% MF, contraction: SEM, 1.8%–3%
Koppenhaver et al,[42] 2009[a]	30, NSLBP	SonoSite Titan US unit in B-mode with a 60-mm, 2–5-MHz curvilinear array transducer	TrA and MF thickness at rest and contraction	Inter-rater, intra-rater (1–3 d)	Inter-rater $ICC_{2,2}$ TrA and MF: 0.80–0.94 Intra-rater $ICC_{3,2}$ TrA and MF: 0.93–0.99	Inter-rater TrA and MF: SEM, 0.2–2.1; MDC, 0.7-5.8 Intra-rater TrA and MF: SEM, 0.1–1.1; MDC, 0.4–3.1
Jhu et al,[31] 2010[a]	H	HDI 5000 US unit (Philips Healthcare, Bothell, WA) in B-mode with a 5–12-MHz transducer, automatically adjusted by the scanning depth	TrA thickness at rest and Contraction	Intra-rater (2 d)	$ICC_{3,1}$ TrA: 0.77–0.98	TrA: SEM, 0.2–0.3 MDC, 0.5–0.8
Pinto et al,[44] 2011	20: 10, H; 10, CLBP	Sonoline SL1 US unit (Siemens Ltda, São Paulo, Brazil) with a 10-cm, 7.5-MHz linear array transducer	TrA and IO thickness at rest and contraction	Intra-rater (7 d)	$ICC_{3,1}$ TrA:0.89 IO: 0.75	TRA: SEM, 2.34% IO: SEM, 1.46%

(Continued)

Table 3. Continued

Authors, y	Sample Size, Status	US Setting	Muscles, Dimensions, Status	Type of Reliability (Interval)	Reliability Estimate	Precision Estimate
Pinto et al,[45] 2011	8, CLBP; 7, H	Siemens Sonoline SL1 US unit with a 10-cm, 7.5-MHz linear array transducer	TrA, IO, and EO thickness at rest and contraction	Intra-rater	$ICC_{3,1}$ TrA: 0.78–0.97 IO: 0.69–0.77 EO: 0.11–0.83	TrA: SEM, 5.7% IO: SEM, 5.0% EO: SEM, 4.6%
Rasouli et al,[47] 2011	10, H	ES500 US unit, (Ultrasonix, Richmond, British Columbia, Canada) in B-mode with a 7.5-MHz linear array transducer	TrA and IO	Intra-rater (same d)	$ICC_{3,1}$ TrA and IO: >0.86	
Teyhen et al,[32] 2011[a]	21, H	Portable US unit (SonoSite Titan) in B-mode with a 60-mm, 5-MHz curvilinear array transducer	IO, TrA, and MF thickness, RA CSA and thickness at rest and contraction	Inter-rater (same d)	$ICC_{1,3}$ TrA, IO, and MF: 0.86–0.94	TrA, IO, and MF: SEM, 0.04–0.70; MDC, 010–0.46
Watson et al,[33] 2011[a]	6, H	Portable US unit (InNovaSound USB; Direct Medical Systems, Pleasanton, CA) in B-mode with 3.0-cm, 7.5-MHz linear array transducer	TrA at rest and contraction	Intra-rater, inter-rater	$ICC_{2,1}$ Inter-rater: TrA: rest, 0.71–0.83 Intra-rater: TrA: rest, 0.97–0.99	Inter-rater: SEM, 0.16–0.30; MDC, 0.44–0.82 Intra-rater: SEM, 0.05–0.07; MDC, 0.14–0.20
Gnat et al,[35] 2012[a]	10, H	DP 6600 US unit (Mindray, Mahwah, NJ) in B-mode with a 75L38EA 5-MHz linear array transducer	TrA thickness at rest and contraction	Inter-rater, intra-rater (start, 2 and 4 d)	Inter-rater: $ICC_{2,1}$, $ICC_{2,2}$, $ICC_{2,3}$: 0.86–0.98 Intra-rater: $ICC_{2,1}$, $ICC_{2,2}$, $ICC_{2,3}$: 0.86–0.97	Inter-rater: SEM, 0.18–0.48; MDC, 0.60–1.34 Intra-rater: SEM, 0.22–0.48; MDC, 0.60–1.34
McPherson and Watson,[34] 2012[a]	13, H	InNovaSound USB US unit in B-mode with a 3.0-cm 7.5-MHz linear array transducer	TrA thickness at contraction	Intra-rater	$ICC_{2,k}$ TrA: 0.62–0.93	TrA: SEM, 0.24–0.56; MDC, NA
Arab et al,[36] 2013[a]	20, H, CLBP	Ultrasonix ES500 US unit in B-mode with a 7.5-MHz linear array transducer	TrA, IO, EO, and RA thickness	Intra-rater (same d [H and CLBP] 1 wk [H])	TrA, IO, EO, and RA within-d $ICC_{3,1}$ H: 0.88–0.95 CLBP: 0.89–0.94 Between-d $ICC_{3,1}$: 0.85–0.94	Within-d SEM: H, 0.19–0.55; CLBP, 0.25–0.93 MDC: H, 0.52–1.51; CLBP, 0.69–3.50 Between-d SEM: 0.2–0.71; MDC, 0.55–1.96
Larivière et al,[51] 2013	30: 15, H; 15, CLBP	Philips HD11 1.0.6 US unit in B-mode with a 6.5-cm, 2–5-MHz curvilinear array transducer	MF thickness at rest and contraction	Inter-rater, intra-rater (7–14 d)	Inter-rater ICC MF: 0.68–0.79 Intra-rater ICC MF: 0.89–0.94	Inter-rater SEM MF: 3.1–3.8 Intra-rater SEM MF: 1.5–2.00

(Continued)

Table 3. Continued

Authors, y	Sample Size, Status	US Setting	Muscles, Dimensions, Status	Type of Reliability (Interval)	Reliability Estimate	Precision Estimate
Djordjevic et al,[38] 2014[a]	42, H; 56, LBP	Nemio SSA-550 A US unit (Toshiba Medical Systems, Tokyo, Japan) in B-mode with a 3.75-MHz curvilinear array transducer	TrA and MF thickness at rest and contraction	Inter-rater, intra-rater (once/d on 3 consecutive d)	TrA and MFL: Inter-rater $ICC_{2,1}$ and $ICC_{2,3}$ H, 0.93–1.00; LBP, 0.71–0.87 Intra-rater $ICC_{2,3}$ H, 0.93–1.00; LBP, 0.95–1.00	TrA and MF Inter-rater SEM: H, 0.11–0.63; LBP, 0.12–0.64 Inter-rater MDC: H, 0.29–1.73; LBP, 0.33–1.76, Intra-rater SEM: H, 0.14–0.91; LBP, 0.16–0.88 Intra-rater MDC: H, 0.39–2.52; LBP, 0.43–2.35
Nabavi et al,[50] 2014	15, H	Ultrasonix ES500 US unit with 7.5-MHz linear and 3.5-MHz curvilinear array transducers	LMF thickness and CSA, TrA thickness	Intra-rater (same d and 1 wk)	Within-d Thickness: TrA, 0.80–0.84; LMF, 0.89–0.94 CSA: LMF, 0.68–0.84 Between-d Thickness: TrA, 0.74–0.86; LMF, 0.89–0.92 CSA: LMF, 0.74–0.88	Within-d SEM Thickness: TrA, 0.30–0.41; LMF, 1.19–1.74 CSA: LMF, 0.36–0.74 Between-d SEM Thickness: TrA, 0.28–0.47; LMF, 1.35–1.53 CSA: LMF, 0.33–0.53
Tahan et al,[37] 2014[a]	21, H, LBP	Ultrasonix ES500 US unit in B-mode with a 7.5-MHz linear array transducer	TrA and IO thickness at rest and contraction	Intra-rater (same d)	TrA and IO: $ICC_{3,1}$ H, 0.77–0.92; LBP, 0.74–0.96	TrA and IO: SEM H, 0.2–0.4; LBP, 0.2–0.8
Hoppes et al,[39] 2015[a]	33, H	SonoSite M-Turbo US unit in B-mode with a 60-mm, 2–5-MHz curvilinear array transducer	TrA and IO thickness at rest and contraction	Inter-rater, intra-rater	TrA and IO Inter-rater $ICC_{2,1}$: 0.39–0.79 Intra-rater $ICC_{3,3}$: 0.90–0.98	TrA and IO: Inter-rater SEM, 0.07-0.16; MDC, 0.49–1.41 Intra-rater SEM, 0.03–0.07; MDC, 0.35–1.10
Hosseinifar et al,[48] 2015[a]	15, H	My LabX Vision 50 US unit (Esaote SpA, Genoa, Italy) in B-mode with a 7.5-MHz curvilinear array transducer	MF thickness at rest and contraction, MF CSA at rest	Intra-rater (2 d)	ICC MF thickness: 0.66–0.87 MF CSA: 0.74–0.91	MF thickness: SEM, 0.63–1.05 MF CSA: SEM, 5.35–6.26
Sions et al,[40] 2015[a]	31, CLBP		MF thickness at rest and contraction	Inter-rater, intra-rater (2–9 d)	Inter-rater $ICC_{2,3}$ Within-d: 0.82–0.98	Within-d inter-rater: SEM, 0.12–0.40;

(Continued)

Table 3. Continued

Authors, y	Sample Size, Status	US Setting	Muscles, Dimensions, Status	Type of Reliability (Interval)	Reliability Estimate	Precision Estimate
		Esaote MyLab 25 portable US unit in B-mode with a 3.5–7.0-MHz curvilinear array transducer			Between-d: 0.72–0.79 Intra-rater $ICC_{3,3}$ Between-d: 0.91–0.93	MDC, 0.34–1.10 Between-d inter-rater: SEM, 0.41–0.46; MDC, 1.13–1.26 Between-d intra-rater: SEM, 0.21–0.24; MDC: 0.60-0.66
Ehsani et al,[29] 2016[a]	46: 23, H; 23, CLBP	HS-2100V US Unit (Honda Electronics, Aichi, Japan) in B-mode with a 7.5-MHz linear array transducer	TrA, IO, and EO thickness	Intra-rater (3–5 d)	TrA, IO, and EO Within-d $ICC_{1,4}$ H: 0.0.67–0.98 CLBP: 0.93–0.98 Between-d $ICC_{2,1}$ H: 0.63–0.99 CLBP: 0.90–0.99	TrA, IO, and EO: Within-d H: SEM, 0.04–1.05; MDC, 0.08–1.32 CLBP: SEM, 0.03–0.89; MDC, 0.05–1.75 Between-d H: SEM, 0.04–1.05; MDC, 0.08–2.06 CLBP: SEM, 0.03–0.89; MDC, 0.05–1.75
Nagai et al,[49] 2016	20, H (reliability in 11 of 20)	ProSound 6 US unit (Aloka Co, Ltd, Tokyo, Japan) in B-mode with a 7.5-MHz linear array transducer	TrA and IO thickness at contraction	Intra-rater (7–14 d)	$ICC_{1,1}$ TrA: 0.76–0.90 IO: 0.84–0.91	
Wilson et al,[41] 2016[a]	92, H (82%), LBP (18%)	Philips HDI 5000 US unit in B-mode with a handheld curved array transducer	RA and EO thickness, MF CSA (L2–5) at rest, IO, TrA, and MF thickness at rest and contraction	Inter-rater, intra-rater (7–10 d)	Inter-rater $ICC_{3,1}$ TrA, IO, EO, and MF thickness: 0.86–0.99 MF CSA (L5): 0.92 Intra-rater $ICC_{3,1}$ TrA, IO, EO, and MF thickness: 0.95–0.99 MF CSA (L2-L5): 0.93–0.97	Inter-rater TrA, IO, EO, and MF thickness: SEM, 0.02–0.20; MDC, 0.04–0.55 MF CSA (L5): SEM, 0.51; MDC, 1.40 Intra-rater TrA, IO, EO, and MF thickness: SEM, 0.02–0.19; MDC, 0.05–0.53 MF CSA (L2–L5): SEM, 0.20–0.37; MDC, 0.56–1.03

(Continued)

Table 3. Continued

Authors, y	Sample Size, Status	US Setting	Muscles, Dimensions, Status	Type of Reliability (Interval)	Reliability Estimate	Precision Estimate
Aboufazeli and Afshar-Mohajer,[52] 2017	20: 10, H; 10, LBP	LOGIQ S6 US unit (GE Healthcare, Milwaukee, WI) in B-mode with a 7.5-MHz linear array transducer	TrA, IO, and EO thickness at contraction	Intra-rater (same d and 5 d)	Within-d $ICC_{3,1}$ TrA and IO: 0.72–0.87 EO: 0.42–0.82 Between-d TrA and IO: 0.62–0.80 EO: 0.33–0.87	Within-d: SEM, 0.16–0.68; MDC, 0.44–1.87 Between-d: SEM, 0.35–0.78; MDC, 0.96–2.15

CLBP indicates chronic LBP; CSA, cross-sectional area; EO, external oblique; H, healthy; IO, internal oblique; LMF, lumbar multifidus; MDC, minimal detectable change; MF, multifidus; NA, not available; NSLBP, nonspecific LBP; RA, rectus abdominis; SEM, standard error of measurement; and TrA, transversus abdominis.
[a]High-quality studies.

the included studies ranged from 10% to 90%. Also, 16 of 26 reviewed studies (marked [a] in Table 3) obtained a score of 50% or greater and were deemed to be high-quality articles.[27–42]

Intraclass correlation coefficients of 0.75 or greater are assumed to indicate an acceptable level of reliability.[25] The intra-rater reliability of trunk stabilizer muscles was evaluated in all studies the except study by Teyhen et al.[43] Among high-quality studies, the intra-rater reliability was generally in an acceptable range (ICC, $\geq$ 0.70). Of high-quality studies, McPherson and Watson[34] (ICC, 0.62–0.93) and Ehsani et al[29] (ICC, 0.63–0.99) reported intra-rater reliability ICCs that were less than 0.70. The ICCs reported for inter-rater reliability in high-quality studies were greater than 0.70 (0.71–1.00),[32,38,42] except for Hoppes et al[39] (0.39–0.79). Also, among low-quality studies, reported ICCs for intra- and inter-rater reliability ranged from 0.26 to 0.99 and 0.68 to 0.97, respectively.

As in 4 studies, the reliability of rehabilitative US imaging was reported within the main study; no data were reported on the sex and body mass index of the patients.[27,43–45] Among the included articles, 12 studies evaluated the reliability of rehabilitative US imaging in healthy individuals.[31–35,39,43,46–50] Five studies evaluated the reliability in patients with LBP,[27,28,30,40,42] and 9 studies reported the reliability for both healthy individuals and those with LBP.[29,36–38,41,44,45,51,52] The range of inter- and intra-rater reliability for each muscle is summarized in Table 5.

Discussion

This study was designed to systematically review published articles from June 2007 to August 2017 to determine whether rehabilitative US imaging is a reliable method for measuring abdominal and lumbar multifidus muscle dimensions. Results of the reviewed studies indicated that the reliability of rehabilitative US imaging ranged from poor to excellent, and the overall reported reliabilities were in an acceptable range. Lower ICC values were commonly related to unrepresentative samples of participants or examiners, and replication of the assessment procedure was impossible.[45,52,53]

As presented in Table 4, 8 of the 9 low-quality articles did not use representative samples of participants and examiners.[43–47,49,50,52] Furthermore, in all low-quality studies, neither the examiners nor the participants were blinded.

One low-quality study reported a very wide range of intra-rater reliability for application of rehabilitative US to assess the external oblique muscle (ICC: 0.11–0.83).[45] Lower ICC values were reported in 2 high-quality studies, which may be attributed to different tasks, loads, and positions while rehabilitative US imaging was performed.[29,34]

Table 4. Methodological Quality Scores of Included Studies

Authors	y	External Validity			Internal Validity					Statistical Method		
		Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	% Y
Djordjevic et al[38]	2014	Y	Y	Y	Y	Y	Y	Y	?	Y	Y	90
Koppenhaver et al[42]	2009	Y	Y	Y	Y	Y	?	Y	Y	Y	Y	90
Ehsani et al[29]	2016	Y	Y	N	Y	Y	Y	Y	?	Y	Y	80
Sions et al[40]	2015	Y	Y	Y	?	Y	?	Y	Y	Y	Y	80
Wilson et al[41]	2016	N	Y	Y	Y	Y	?	Y	?	Y	Y	70
Hoppes et al[39]	2015	N	Y	N	?	Y	Y	Y	Y	Y	Y	70
Gnat et al[35]	2012	N	Y	Y	Y	Y	?	?	Y	Y	Y	70
Koppenhaver et al[30]	2009	Y	Y	Y	Y	Y	?	Y	?	Y	N	70
Tahan et al[37]	2014	Y	N	Y	Y	Y	?	Y	?	Y	?	60
Teyhen et al[32]	2011	N	Y	Y	Y	?	?	Y	?	Y	Y	60
Watson et al[33]	2011	N	Y	N	?	Y	Y	Y	?	Y	Y	60
Costa et al[28]	2009	Y	N	Y	Y	Y	?	?	Y	Y	N	60
Kiesel et al[27]	2007	Y	?	N	Y	Y	?	Y	?	Y	Y	60
Arab et al[36]	2013	?	N	N	Y	Y	?	Y	?	Y	Y	50
McPherson and Watson[34]	2012	N	N	N	?	Y	Y	Y	Y	Y	N	50
Jhu et al[31]	2010	N	N	Y	Y	Y	?	?	?	Y	Y	50
Pinto et al[44]	2011	N	N	Y	Y	Y	?	?	?	Y	N	40
Wallwork et al[46]	2007	N	N	Y	Y	Y	?	?	?	Y	?	40
Aboufazeli et al[52]	2017	N	N	N	Y	?	?	?	?	Y	Y	30
Nabavi et al[50]	2014	N	N	Y	Y	Y	?	?	?	N	N	30
Larivière et al[51]	2013	Y	Y	N	?	Y	?	?	?	?	N	30
Rasouli et al[47]	2011	N	N	N	Y	Y	?	?	?	Y	N	30
Teyhen et al[43]	2008	N	N	Y	Y	?	?	?	?	Y	?	30
Nagai et al[49]	2016	N	N	N	Y	Y	?	?	?	?	N	20
Hosseinifar et al[48]	2015	N	N	Y	Y	?	?	?	?	?	?	20
Pinto et al[45]	2011	N	N	N	?	N	?	?	?	Y	N	10

N indicates no; Y, yes; and ?, unclear.

Table 5. Range of Reported ICCs for Trunk Stabilizer Muscles

Muscle	Status	Healthy		LBP	
		Inter-rater	Intra-rater	Inter-rater	Intra-rater
TrA	Rest	0.39–0.99	0.74–0.99	0.71–0.99	0.63–0.99
	Contraction	0.46–0.98	0.62–0.98	0.73–0.98	0.41–1.00
IO	Rest	0.64–0.99	0.67–0.99	0.99	0.69–0.99
	Contraction	0.76–0.99	0.62–0.99	0.99	0.62–0.99
EO	Rest	0.98	0.11–0.98	0.98	0.11–0.99
	Contraction		0.33–0.98	NA	0.33–0.99
MF CSA	Rest	0.92	0.74–0.97	0.92	0.93–0.97
	Contraction				
MF thickness	Rest	0.68–1.00	0.66–1.00	0.78–0.98	0.95–1.00
	Contraction	0.70–0.97	0.64–1.00	0.70–0.98	0.86–1.00

CSA indicates cross-sectional area; EO, external oblique; IO, internal oblique; MF, multifidus; NA, not available; and TrA, transversus abdominis.

Therefore, it is suggested that future reliability studies should focus on representative samples of participants and examiners to obtain higher ICCs. Also, different parameters related to the replicability of the study (including scanning point, amount of contact pressure, muscle state, transducer position, and participant's starting position) should be kept constant.

Among the included studies that measured abdominal muscle thickness, 14 studies[28,29,31,33–37,44,45,47,49,50,52] used linear transducers, and 8 studies[27,30,32,38,39,41–43] used curvilinear transducers with frequencies of 5–10 MHz (commonly 7.5) and 2–5 MHz, respectively. Also, curvilinear transducers were applied in all studies that measured lumbar multifidus thickness or cross-sectional area,[27,30,32,38,40–42,46,48,51] except 1 study,[50] that used a linear transducer. The most commonly used frequencies for measurement of the lumbar multifidus were 5 and 7.5 MHz.[27,30,32,42,46,48,50]

In most studies US images were taken at the end of expiration to control the influence of respiration on trunk muscle dimensions. Furthermore, the most popular positions for measurement of abdominal and multifidus muscles were supine or supine hook-lying and prone, respectively. All included studies that measured lumbar multifidus muscle thickness used curvilinear transducers laterally and angled medially to the lumbar spinal process to capture an image of the zygapophyseal joint. Then, the distance between the dorsal edge of the joint and the hyperechoic thoracolumbar fascia was considered the thickness of the muscle.[27,30,32,38,40–42,46,48,51]

As presented in Table 5, no published study was found to evaluate inter- and intra-rater reliability of the lumbar multifidus during contraction. Therefore, future studies are recommended to evaluate the intra-rater reliability of rehabilitative US imaging for measurement of lumbar multifidus muscle dimensions during contraction. In 16 included high-quality studies, the reported ICCs for intra-rater reliability ranged from 0.62 to 1.00 in healthy individuals and patients with LBP. Furthermore, the inter-rater reliability reported in high-quality studies ranged from 0.71 to 1.00, except for the study by Hoppes et al[39] (transversus abdominis, 0.39–0.72; internal oblique, 0.52–0.79). This finding might be attributed to images taken by a novice rater.[39]

In 14 studies, appropriate estimates of both reliability and precision were reported,[27,29,31–33,35,36,38–42,52,53] and in 6 studies, representative samples of participants and examiners were reported.[29,30,38,40,42,51] Some studies mentioned the results for rehabilitative US reliability inside a randomized controlled trial or other main articles. Therefore, some demographic information, such as age, sex, body mass index, and muscle dimension, were not reported.[27,29,33,43–45,47,49]

The results of this study are in agreement with a previous systematic review conducted on the reliability of rehabilitative US imaging in the assessment of abdominal and lumbar trunk muscles by Hebert et al.[23] In general, our systematic review indicates that rehabilitative US imaging has acceptable reliability for evaluating abdominal and lumbar trunk muscles at rest and during contraction in individuals with and without LBP, particularly when measurements are taken by experienced examiners. Having a good amount of reliability is clinically valuable, as rehabilitative US imaging is a safe and cost-effective method that can be reliably used for measurement of trunk muscle dimensions compared to other methods, including magnetic resonance imaging and computed tomography.[20] It can be applied both for the assessment of muscle morphologic characteristics in different conditions and to monitor the effectiveness of interventions in a clinical setting. It may be also useful for differentiating people with a certain condition and those at risk from healthy people. Furthermore, using standard and consistent methods for parameters such as the participant's position, determination of the scanning point, and the amount of contact pressure from the transducer in different sessions may improve the results for rater reliability.

There were some limitations in this study. First, only English-language articles were included. Second, the syntax for the search in each database was not defined, and the search was performed by key words. Finally, a meta-analysis was not performed in this study. However, more systematic reviews are needed to evaluate the reliability of rehabilitative US imaging with meta-analyses and fewer limitations.

In conclusion, according to the results of this systematic review, rehabilitative US imaging is a reliable method for evaluating deep abdominal and lumbar trunk muscles at rest and during contraction in healthy individuals and patients with LBP. Because of the lack of methodological consistency between studies using rehabilitative US imaging to measure lumbar and abdominal trunk muscles, different reliability values were reported in previously conducted studies of different quality.

References

1. Pourahmadi MR, Taghipour M, Jannati E, Mohseni-Bandpei MA, Takamjani IE, Rajabzadeh F. Reliability and validity of an iPhone® application for the measurement of lumbar spine flexion and extension range of motion. *Peer J* 2016; 4:e2355.
2. Manchikanti L, Singh V, Falco FJ, Benyamin RM, Hirsch JA. Epidemiology of low back pain in adults. *Neuromodulation* 2014; 17(suppl 2):3–10.
3. Hoy D, Bain C, Williams G, et al. A systematic review of the global prevalence of low back pain. *Arthritis Rheum* 2012; 64:2028–2037.
4. Mohseni-Bandpei MA, Rahmani N, Halimi F, Farooq MN. The prevalence of low back pain in Iranian dentists: an epidemiological study. *Pak J Med Sci* 2017; 33:280–284.
5. Mohseni-Bandpei MA, Ahmad-Shirvani M, Golbabaei N, Behtash H, Shahinfar Z, Fernández-de-las-Peñas C. Prevalence and risk factors associated with low back pain in Iranian surgeons. *J Manipulative Physiol Ther* 2011; 34:362–370.
6. Panjabi MM. The stabilizing system of the spine, part I: function, dysfunction, adaptation, and enhancement. *Clin Spine Surg* 1992; 5:383–389.
7. Franke J, Hesse T, Tournier C, et al. Morphological changes of the multifidus muscle in patients with symptomatic lumbar disc herniation: clinical article. *J Neurosurg Spine* 2009; 11:710–714.
8. Kim WH, Lee SH, Lee DY. Changes in the cross-sectional area of multifidus and psoas in unilateral sciatica caused by lumbar disc herniation. *J Korean Neurosurg Soc* 2011; 50:201–204.
9. Ferreira PH, Ferreira ML, Hodges PW. Changes in recruitment of the abdominal muscles in people with low back pain: ultrasound measurement of muscle activity. *Spine* 2004; 29:2560–2566.
10. Ghamkhar L, Emami M, Mohseni-Bandpei MA, Behtash H. Application of rehabilitative ultrasound in the assessment of low back pain: a literature review. *J Bodyw Mov Ther* 2011; 15:465–477.
11. Hodges P, Holm AK, Hansson T, Holm S. Rapid atrophy of the lumbar multifidus follows experimental disc or nerve root injury. *Spine* 2006; 31:2926–2933.
12. Hides J, Stokes M, Saide M, Jull G, Cooper D. Evidence of lumbar multifidus muscle wasting ipsilateral to symptoms in patients with acute/subacute low back pain. *Spine* 1994; 19:165–172.
13. Leinonen V, Kankaanpää M, Luukkonen M, et al. Lumbar paraspinal muscle function, perception of lumbar position, and postural control in disc herniation-related back pain. *Spine* 2003; 28:842–848.
14. Mohseni-Bandpei MA, Rahmani N, Majdoleslam B, Abdollahi I, Ali SS, Ahmad A. Reliability of surface electromyography in the assessment of paraspinal muscle fatigue: an updated systematic review. *J Manipulative Physiol Ther* 2014; 37:510–521.
15. Kader D, Wardlaw D, Smith F. Correlation between the MRI changes in the lumbar multifidus muscles and leg pain. *Clin Radiol* 2000; 55:145–149.
16. Mattila M, Hurme M, Alaranta H, et al. The multifidus muscle in patients with lumbar disc herniation: a histochemical and morphometric analysis of intraoperative biopsies. *Spine* 1986; 11:732–738.
17. Kamaz M, Kiresi D, Oguz H, Emlik D, Levendoglu F. CT measurement of trunk muscle areas in patients with chronic low back pain. *Diagn Interv Radiol* 2007; 13:144–148.
18. Taghipour M, Rajabzadeh F, Arab AM. The relationship between the side-lying endurance test and ultrasound thickness measurement for abdominal muscles. *Jentashapir J Health Res* 2015; 6:e29206.
19. Rajabzadeh F, Taghipour M, Pourahmadi MR, Arab AM. Relationship between flexion-rotation test and ultrasound thickness measurement of the abdominal muscles. *J Back Musculoskelet Rehabil* 2016; 29: 1–6.
20. Whittaker JL, Teyhen DS, Elliott JM, et al. Rehabilitative ultrasound imaging: understanding the technology and its applications. *J Orthop Sports Phys Ther* 2007; 37:434–449.
21. Bartko JJ, Carpenter WT. On the methods and theory of reliability. *J Nerv Ment Dis* 1976; 163:307–317.
22. Peter JP. Reliability: a review of psychometric basics and recent marketing practices. *J Marketing Res* 1979; 16:6–17.
23. Hebert JJ, Koppenhaver SL, Parent EC, Fritz JM. A systematic review of the reliability of rehabilitative ultrasound imaging for the quantitative assessment of the abdominal and lumbar trunk muscles. *Spine* 2009; 34:E848–E856.
24. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med* 2009; 6:e1000100.
25. van Trijffel E, Anderegg Q, Bossuyt P, Lucas C. Inter-examiner reliability of passive assessment of intervertebral motion in the cervical and lumbar spine: a systematic review. *Man Ther* 2005; 10:256–269.
26. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960; 20:37–46.
27. Kiesel KB, Underwood FB, Mattacola CG, Nitz AJ, Malone TR. A comparison of select trunk muscle thickness change between subjects with low back pain classified in the treatment-based classification system and asymptomatic controls. *J Orthop Sports Phys Ther* 2007; 37: 596–607.
28. Costa LOP, Maher CG, Latimer J, Hodges PW, Shirley D. An investigation of the reproducibility of ultrasound measures of abdominal muscle activation in patients with chronic non-specific low back pain. *Eur Spine J* 2009; 18:1059–1065.
29. Ehsani F, Arab AM, Jaberzadeh S, Salavati M. Ultrasound measurement of deep and superficial abdominal muscles thickness during standing postural tasks in participants with and without chronic low back pain. *Man Ther* 2016; 23:98–105.
30. Koppenhaver SL, Parent EC, Teyhen DS, Hebert JJ, Fritz JM. The effect of averaging multiple trials on measurement error during ultrasound imaging of transversus abdominis and lumbar multifidus muscles in individuals with low back pain. *J Orthop Sports Phys Ther* 2009; 39:604–611.
31. Jhu JL, Chai HM, Jan MH, Wang CL, Shau YW, Wang SF. Reliability and relationship between two measurements of transversus abdominis

dimension taken during an abdominal drawing-in maneuver using a novel approach of ultrasound imaging. *J Orthop Sports Phys Ther* 2010; 40:826–832.

32. Teyhen DS, George SZ, Dugan JL, Williamson J, Neilson BD, Childs JD. Inter-rater reliability of ultrasound imaging of the trunk musculature among novice raters. *J Ultrasound Med* 2011; 30:347–356.
33. Watson T, McPherson S, Fleeman S. Ultrasound measurement of transversus abdominis during loaded, functional tasks in asymptomatic individuals: rater reliability. *PM R* 2011; 3:697–705.
34. McPherson SL, Watson T. Reproducibility of ultrasound measurement of transversus abdominis during loaded, functional tasks in asymptomatic young adults. *PM R* 2012; 4:402–412.
35. Gnat R, Saulicz E, Miądowicz B. Reliability of real-time ultrasound measurement of transversus abdominis thickness in healthy trained subjects. *Eur Spine J* 2012; 21:1508–1515.
36. Arab AM, Rasouli O, Amiri M, Tahan N. Reliability of ultrasound measurement of automatic activity of the abdominal muscle in participants with and without chronic low back pain. *Chiropr Man Ther* 2013; 21:37.
37. Tahan N, Rasouli O, Arab AM, Khademi K, Samani EN. Reliability of the ultrasound measurements of abdominal muscles activity when activated with and without pelvic floor muscles contraction. *J Back Musculoskelet Rehabil* 2014; 27:339–347.
38. Djordjevic O, Djordjevic A, Konstantinovic L. Interrater and intrarater reliability of transverse abdominal and lumbar multifidus muscle thickness in subjects with and without low back pain. *J Orthop Sports Phys Ther* 2014; 44:979–988.
39. Hoppes CW, Sperier AD, Hopkins CF, et al. Ultrasound imaging measurement of the transversus abdominis in supine, standing, and under loading: a reliability study of novice examiners. *Int J Sports Phys Ther* 2015; 10:910–917.
40. Sions JM, Velasco TO, Teyhen DS, Hicks GE. Reliability of ultrasound imaging for the assessment of lumbar multifidi thickness in older adults with chronic low back pain. *J Geriatr Phys Ther* 2015; 38:33–39.
41. Wilson A, Hides JA, Blizzard L, et al. Measuring ultrasound images of abdominal and lumbar multifidus muscles in older adults: a reliability study. *Man Ther* 2016; 23:114–119.
42. Koppenhaver SL, Hebert JJ, Fritz JM, Parent EC, Teyhen DS, Magel JS. Reliability of rehabilitative ultrasound imaging of the transversus abdominis and lumbar multifidus muscles. *Arch Phys Med Rehabil* 2009; 90:87–94.
43. Teyhen DS, Rieger JL, Westrick RB, Miller AC, Molloy JM, Childs JD. Changes in deep abdominal muscle thickness during common trunk-strengthening exercises using ultrasound imaging. *J Orthop Sports Phys Ther* 2008; 38:596–605.
44. Pinto RZ, Ferreira PH, Franco MR, et al. The effect of lumbar posture on abdominal muscle thickness during an isometric leg task in people with and without non-specific low back pain. *Man Ther* 2011; 16: 578–584.
45. Pinto RZ, Ferreira PH, Franco MR, et al. Effect of two lumbar spine postures on transversus abdominis muscle thickness during a voluntary contraction in people with and without low back pain. *J Manipulative Physiol Ther* 2011; 34:164–172.
46. Wallwork TL, Hides JA, Stanton WR. Intrarater and interrater reliability of assessment of lumbar multifidus muscle thickness using rehabilitative ultrasound imaging. *J Orthop Sports Phys Ther* 2007; 37:608–612.
47. Rasouli O, Arab AM, Amiri M, Jaberzadeh S. Ultrasound measurement of deep abdominal muscle activity in sitting positions with different stability levels in subjects with and without chronic low back pain. *Man Ther* 2011; 16:388–393.
48. Hosseinifar M, Akbari A, Ghiasi F. Intra-rater reliability of rehabilitative ultrasound imaging for multifidus muscles thickness and cross section area in healthy subjects. *Glob J Health Sci* 2015; 7:354–361.
49. Nagai H, Akasaka K, Otsudo T, Sawada Y, Okubo Y. Deep abdominal muscle thickness measured under sitting conditions during different stability tasks. *J Phys Ther Sci* 2016; 28:900–905.
50. Nabavi N, Mosallanezhad Z, Haghighatkhah HR, Bandpeid MAM. Reliability of rehabilitative ultrasonography to measure transverse abdominis and multifidus muscle dimensions. *Iran J Radiol* 2014; 11: e21008.
51. Larivière C, Gagnon D, De Oliveira E, Henry SM, Mecheri H, Dumas JP. Ultrasound measures of the lumbar multifidus: effect of task and transducer position on reliability. *PM R* 2013; 5:678–687.
52. Aboufazeli M, Afshar-Mohajer N. Within-day and between-day reliability of thickness measurements of abdominal muscles using ultrasound during abdominal hollowing and bracing maneuvers. *J Bodyw Mov Ther* 2018; 22:122–128.
53. Mannion AF, Pulkovski N, Gubler D, et al. Muscle thickness changes during abdominal hollowing: an assessment of between-day measurement error in controls and patients with chronic low back pain. *Eur Spine J* 2008; 17:494–501.