


Reliability of lumbar multifidus and iliocostalis lumborum thickness and echogenicity measurements using ultrasound imaging

Joshua Farragher¹ , Adrian Pranata², Doa El-Ansary^{2,3,4}, Selina Parry⁵, Gavin Williams^{5,6}, Colin Royse^{3,7,8}, Alistair Royse³, Molly O'Donohue⁹ and Adam Bryant¹

¹Centre for Health Exercise & Sports Medicine, The University of Melbourne, Melbourne, Vic., Australia

²Department of Nursing and Allied Health, Swinburne University of Technology, Hawthorn, Vic., Australia

³Department of Surgery, The University of Melbourne, Melbourne, Vic., Australia

⁴Westmead Private Hospital, Clinical Research Institute, Westmead, NSW, Australia

⁵Department of Physiotherapy, The University of Melbourne, Melbourne, Vic., Australia

⁶Epworth Hospital, Richmond, Vic., Australia

⁷The Royal Melbourne Hospital, Melbourne, Vic., Australia

⁸Outcomes Research Consortium, Cleveland Clinic, Cleveland, OH, USA

⁹Capital Radiology, Frankston, Vic., Australia

Abstract

Purpose: To establish the test–retest and inter-rater reliability of lumbar multifidus (LM) and iliocostalis lumborum (IL) muscle thickness and echogenicity as derived using ultrasound imaging.

Methods: Ultrasound images of the LM and IL were collected from 11 healthy participants on two occasions, 1 week apart, by two independent assessors. Measures of LM and IL thickness and echogenicity were subject to test–retest and inter-rater reliability, which was assessed by calculation of an F statistic, the interclass correlation coefficient (ICC), the standard error of measurement, 95% confidence intervals and Bland–Altman plots. This study was given approval by The University of Melbourne Behavioural and Social Sciences Human Ethics Sub-Committee (ref: 1749845).

Results: Assessors A and B showed good to excellent test–retest reliability for LM thickness (ICC_{3,3} A: 0.89 and B: 0.98), LM echogenicity (ICC_{3,3} A: 0.93 and B: 0.95) and IL echogenicity (ICC_{3,3} A: 0.87 and B: 0.83). Test–retest reliability for IL thickness was poor for Assessor A but excellent for Assessor B. Both assessors demonstrated excellent inter-rater reliability for LM thickness and echogenicity (ICC_{2,3}: 0.79 and 0.94), but poor reliability for IL thickness and echogenicity (ICC_{2,3}: 0.00 and 0.39).

Conclusions: Inter-rater and test–retest reliability was excellent for LM but was less reliable for measures of the IL muscle.

Keywords: imaging, lumbar spine, muscle, reliability, ultrasound.

Correspondence

Joshua Farragher, Centre for Health Exercise & Sports Medicine, The University of Melbourne, Level 7, Alan Gilbert Building, Building 104, 161 Barry Street, Melbourne, Vic. 3010, Australia.

Email: jfarragher@student.unimelb.edu.au

doi: 10.1002/ajum.12273

Introduction

Trunk muscles in the lumbar region provide mechanical stability to the lumbar spine.¹ Clinicians hypothesise that muscular stability and control of the lumbar spine are important for managing and rehabilitating people with low back pain. Ultrasound imaging techniques are used to quantify the cross-sectional area and thickness of lumbar muscles. Lumbar

multifidus (LM) forms part of the deeper transversospinalis group^{2–5} and plays a key role in the stability and function of the lumbar spine.^{6,7} LM-related measures have reported decreases in size, asymmetry and contractility in people with low back pain.^{8–10} Importantly, spontaneous recovery of LM size is generally not displayed after an acute episode of low back pain,¹¹ which may indicate decreased force generating capacity to perform daily tasks such as lifting safely.¹² Aside from the LM, larger lumbar extensor muscles, such as the iliocostalis lumborum (IL), also play important roles in maintaining lumbar posture and generating forces for everyday tasks (e.g. lifting).¹³ The IL muscle (most lateral component of the erector spinae muscles) produces large lumbar extension forces, whilst acting as a stabiliser of the lumbar spine during rotatory movements and a decelerator during lumbar flexion.^{13,14} Therefore, optimum lumbar function is contingent upon both IL and the deeper LM muscles possessing healthy morphological features.

Imaging studies have contributed to a greater understanding of paraspinal muscle morphology and the associated changes in people with low back pain. Specifically, magnetic resonance imaging (MRI) studies examining LM muscle composition have demonstrated a positive association between intra-muscular fat content and low back pain^{15,16} and a negative association with lumbar flexion range of movement.¹⁷ Additionally, generalised lumbar intra-muscular fat infiltration without associated cross-sectional area changes has been observed during dormant periods of recurrent low back pain.¹⁸ However, MRI evaluation is expensive, and challenging to access in routine clinical practice.

Echogenicity analysis of ultrasound images may be utilised to quantify the composition of soft tissue. Specifically, echogenicity refers to the ability to reflect or transmit ultrasound waves in the context of surrounding tissues.¹⁹ Within the thigh region (i.e. quadriceps muscle group), higher ultrasound echogenicity values have been correlated strongly with higher amounts of intra-muscular fat²⁰ and poorer muscle strength and power in healthy elderly^{21,22} and clinical populations.²³ Thus, echogenicity measures may be a cost-effective and feasible clinical alternative to MRI for evaluating muscle composition in the lumbar spine. Ultrasound imaging techniques used to quantify LM muscle size (i.e. cross-sectional area and thickness) have demonstrated high levels of intra- and inter-rater reliability in both healthy individuals^{2,4,5,24–27} and those with low back pain.^{3,4,27} However, there is paucity of evidence evaluating test–retest reliability of ultrasound measures (i.e. thickness and echogenicity) of the LM muscle. Similarly, the inter-rater reliability and test–retest reliability of the morphological data obtained using ultrasound in other lumbar extensor muscles are also lacking.

This study aims to investigate the test–retest and inter-rater reliability of thickness and echogenicity measures of the LM and IL muscle using ultrasound imaging in healthy participants. We hypothesised that test–retest and inter-rater reliability of ultrasound-derived thickness and echogenicity measures would be excellent for the LM muscle. Due to the novelty of

ultrasound-related IL measures, we do not have enough evidence to formulate a hypothesis. The Guidelines for Reporting Reliability and Agreement Studies (GRAAS)²⁸ and the reliability domain of the COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN)²⁹ guidelines were followed in the reporting of this manuscript.

Methods and Materials

Design

This was a prospective observational study involving between-session test–retest and within-session inter-rater reliability study. In the context of this study, test–retest reliability refers to the reliability of ultrasound measures acquired from images taken 1 week apart. Inter-rater reliability refers to the agreement between two different raters on ultrasound measures taken from images of the same muscle at the same time point.

Participants

Eleven healthy participants aged 18–65 years were recruited in April 2018 from the community via personal contacts in Melbourne, Australia. Inclusion criteria were (i) no current low back pain and (ii) no history of low back pain. The exclusion criteria for participation in this study were (i) previous spinal surgery, (ii) spinal trauma, (iii) pregnancy, (iv) inflammatory joint disease, (v) neuromuscular disease, (vi) spinal malignancy and/or (vii) inability to understand English. Participants' demographic data including age, gender, mass and height were collected. All participants provided written informed consent.

Assessors

Ultrasound image acquisition and measurements were performed by two independent assessors: (i) Assessor A – radiographer (medical imaging technologist) with 5 years of experience using ultrasonography and post-graduate degree in diagnostic imaging, (ii) Assessor B – musculoskeletal physiotherapist with 5 years of clinical experience and 10 h of training in musculoskeletal ultrasonography. Both assessors underwent a 2-h training session of the image acquisition protocol involving practical use of the Sonosite Edge machine (Sonosite, Fujifilm, Tokyo, JP) and familiarisation of the functions and settings of the device. Additionally, both assessors took part in a 2-h session of basic education and training around the use Image J version 1.52 (NIH, MD, US) software to perform thickness and echogenicity measures. Training sessions were conducted with both assessors present and involved testing the methodology on people not involved in the study. Both assessors were not previously familiar with using the Sonosite Edge machine or the Image J software.

Image acquisition

Each set of images were collected 1 week apart. Sets of images for all participants were collected by Assessor A then Assessor

B on the same day. Assessors were not present whilst the other assessor was imaging the participants. Image acquisition was performed using a Sonosite Edge Machine and a Linear Transducer (6–13 MHz) and a minimal probe compression approach. The overall gain was set to a moderate brightness level, and the ‘near and far’ gain setting was adjusted to ensure consistent brightness across the field. This setting remained constant for all participants throughout the trial. Participants lay prone with a pillow under their abdomen to reduce lumbar lordosis on a plinth.²⁵ Assessors palpated the lumbar spinous processes and marked the spinous processes of L3, L4 and L5 with a non-permanent marker which was removed between assessors to enable blinding for landmarking. Assessors were able to confirm the accuracy of their superficial landmark identification by ultrasound imaging the parasagittal section.² The multifidus muscle was captured both in the parasagittal (longitudinal) plane and in the transverse (cross-sectional) plane. The parasagittal image was taken by placing the probe longitudinally over the zygapophyseal joint of L4/5 to display the thickness of the LM muscle³⁰ (Figure 1a). The cross-sectional image was taken by placing the probe at the level of the L5 spinous process,³⁰ allowing the lateral border of the spinous process, lamina and multifidus muscle to be visualised (Figure 1b). The IL muscle was also captured in the transverse plane by initially placing the probe adjacent to the L3 spinous process. The probe was moved laterally until the IL muscle was visualised deep to the latissimus dorsi and superficial to the quadratus lumborum (Figure 1c). A total of 6 images, two longitudinal LM images, two cross-sectional LM images and two cross-sectional IL images, were collected for each participant during each test session.

Image processing

All measurements were performed post-image acquisition using the ImageJ software. The measurements of interest were the muscle thickness and echogenicity of the LM and IL muscles. The LM muscle thickness was calculated by measuring the

distance between the inside of the superior fascial border and the tip of the L4/5 zygapophyseal joint (Figure 1a). Measurement of the IL muscle thickness was defined by the distance between the inside of the superior fascial border and the inferior fascial border (Figure 1c). LM and IL echogenicity was measured by (i) performing a circumferential trace of the inner fascial border of each muscle^{30,31} and (ii) performing greyscale histogram analysis of the muscle area of interest. Echogenicity is a measure of acoustic reflectance¹⁹ with ranges varying from 0 to 255: zero represents an inability to reflect ultrasound waves with the image appearing black whereas 255 represents complete reflection of ultrasound waves with the image appearing white.³² The mean echointensity value of the greyscale histogram analysis was calculated to quantitatively evaluate muscle tissue quality.^{33,34} All measures were performed in triplicate with the average value reported.² Images were de-identified by an independent researcher and then randomly assigned to each assessor for image analysis. The assessors were blinded to each other’s results.

Sample size

To look for a minimum ICC value of 0.80 inter-rater reliability, we estimated a priori that a minimum of 8 participants were required for each muscle analysed to achieve 80% power with the significance level set at $p < 0.05$. To ensure that enough images were available for statistical analysis, we included 44 LM and 22 IL images in this study.

Statistical analysis

Test–retest and inter-rater reliability was assessed for absolute agreement using F-statistic, standard error of measurement, 95% confidence interval (95% CI) and intraclass correlation analysis (ICC_{3,3} for test–retest reliability, ICC_{2,3} for inter-rater reliability and ICC_{3,1} for intra-rater reliability; $p < 0.05$ ^{35,36}). ICC was interpreted as follows: poor (0.00–0.25), fair (0.26–0.50), moderate (0.51–0.75) and excellent (0.76–1.00) correlation.³⁷ The Bland–Altman (BA) plots were used to visually

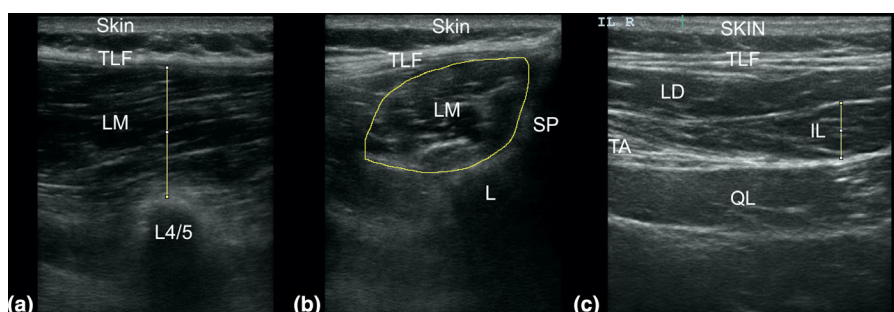


FIGURE 1: Ultrasound image of the left lumbar multifidus (parasagittal) taken at the spinal level of L4/5 (a), the left lumbar multifidus (transverse) taken at the spinal level of L5 (b) and the left iliocostalis lumborum, taken at the spinal level of L3 (c). IL, iliocostalis lumborum; L3, 3rd lumbar vertebrae; L4/5, zygapophyseal joint between L4 and L5; LD, latissimus dorsi; LM, lumbar multifidus; QL, quadratus lumborum; SP, spinous process; TA, transverse abdominus; TLF, thoracolumbar fascia

TABLE 1: Participant characteristics

Characteristic	Participants (n = 11) Mean (SD) or N (%)
Gender (Females, %)	5 (45.5%)
Age (years)	40.6 (13.9)
Height (cm)	173.5 (7.3)
Mass (kg)	82.1 (20.8)
BMI (kg/m ²)	27.0 (5.8)

Abbreviation: BMI, Body Mass Index.

represent the range of agreement. Statistical analyses were performed using SPSS version 23.0 (IBM, Armonk, NY).

Ethical approval

This study was approved by The University of Melbourne Behavioural and Social Sciences Human Ethics Sub-Committee (ref: 1749845) and all participants provided written informed consent and the rights of the subjects were protected.

Results

The participant characteristics are presented in Table 1. The 11 participants (5 females and 6 males) have a mean age of 40.6 years (Table 1). Two participants were considered overweight (25–30 BMI) and 4 obese (>30 BMI), whilst 4 were within the normal weight range (20–25 BMI) and one underweight (<20 BMI).

The ICC calculation pertaining to IL thickness inter-rater reliability produced a negative value, this has been altered to 0.00 as this is not theoretically possible. As such, the lower limits of the 95% CI that produced negative values have also been altered to values of 0.00.

A total of 66 images were recorded and analysed. Thickness and echogenicity measures of the LM conducted by Assessors A and B demonstrated excellent test–retest reliability (Table 2).

TABLE 2: Test–retest reliability for LM and IL muscle thickness and echogenicity

Muscle	Thickness				Echogenicity			
	ICC _{3,3}	95% CI	SEM	F	ICC _{3,3}	95% CI	SEM	F
LM assessor A	0.89	[0.45, 0.97]	0.05	13.63	0.93	[0.72, 0.98]	0.13	12.51
LM assessor B	0.98	[0.94, 0.99]	<0.01	57.34	0.95	[0.82, 0.99]	0.48	19.94
IL assessor A	0.50	[0.00, 0.89]	0.09	1.96	0.87	[0.49, 0.97]	<0.01	6.92
IL assessor B	0.97	[0.90, 0.99]	<0.01	33.72	0.83	[0.37, 0.96]	0.58	5.63

Abbreviations: CI, confidence interval; F, F-statistic; ICC, intraclass correlation coefficient; IL, iliocostalis lumborum; LM, lumbar multifidus; SEM, standard error of measurement.

Measures of IL thickness performed by Assessor A demonstrated fair levels of test–retest reliability, whereas Assessor B demonstrated excellent reliability (Table 2). An *ex-post-facto* analysis that excluded participants with a BMI >30 yielded excellent test–retest reliability for Assessor A regarding IL thickness (ICC_{3,3} = 0.78, 95% CI [0.00, 0.96], n = 7). Echogenicity measurements of the IL displayed excellent test–retest reliability for both assessors (Table 2).

For Assessor A (earlier versus later test), the bias for LM muscle thickness was –0.22 cm and limits of agreement ±0.54 cm (Figure 2a). The bias for IL muscle thickness was 0.18 cm and limits of agreement ±1.55 cm (Figure 2b). For LM and IL echogenicity, the bias was –0.22 and 0.02 greyscale units and limits of agreement were ±15.69 and 22.04 greyscale units, respectively (Figure 2c,d).

For Assessor B (earlier versus later test), the bias for LM muscle thickness was 0.04 cm and limits of agreement ±0.33 cm (Figure 3a). The bias and limits of agreement for IL muscle thickness 0.03 ± 0.36 cm, respectively (Figure 3b). For LM echogenicity, the bias was –3.02 greyscale units and limits of agreement were ±17.07 greyscale units (Figure 3c), and for IL echogenicity, the bias was 1.98 greyscale units and limits of agreement were ±28.71 greyscale units (Figure 3d).

Assessors A and B displayed an excellent level of agreement for measures of LM thickness and echogenicity; however, poor levels of agreement were observed for IL muscle-related thickness and fair levels of agreement for echogenicity measures (Table 3).

For results between assessors (Assessor B versus Assessor A), the bias and limits of agreement for LM and IL muscle thickness were –0.48 ± 0.64 cm and –1.36 ± 1.88 cm, respectively (Figure 4a,b). For LM and IL echogenicity, the bias and limits of agreement were 4.18 ± 13.78 greyscale units and 9.48 ± 39.47 greyscale units, respectively (Figure 4c,d).

Discussion

In accordance with our hypothesis, our results demonstrated excellent test–retest and inter-rater reliability for LM muscle

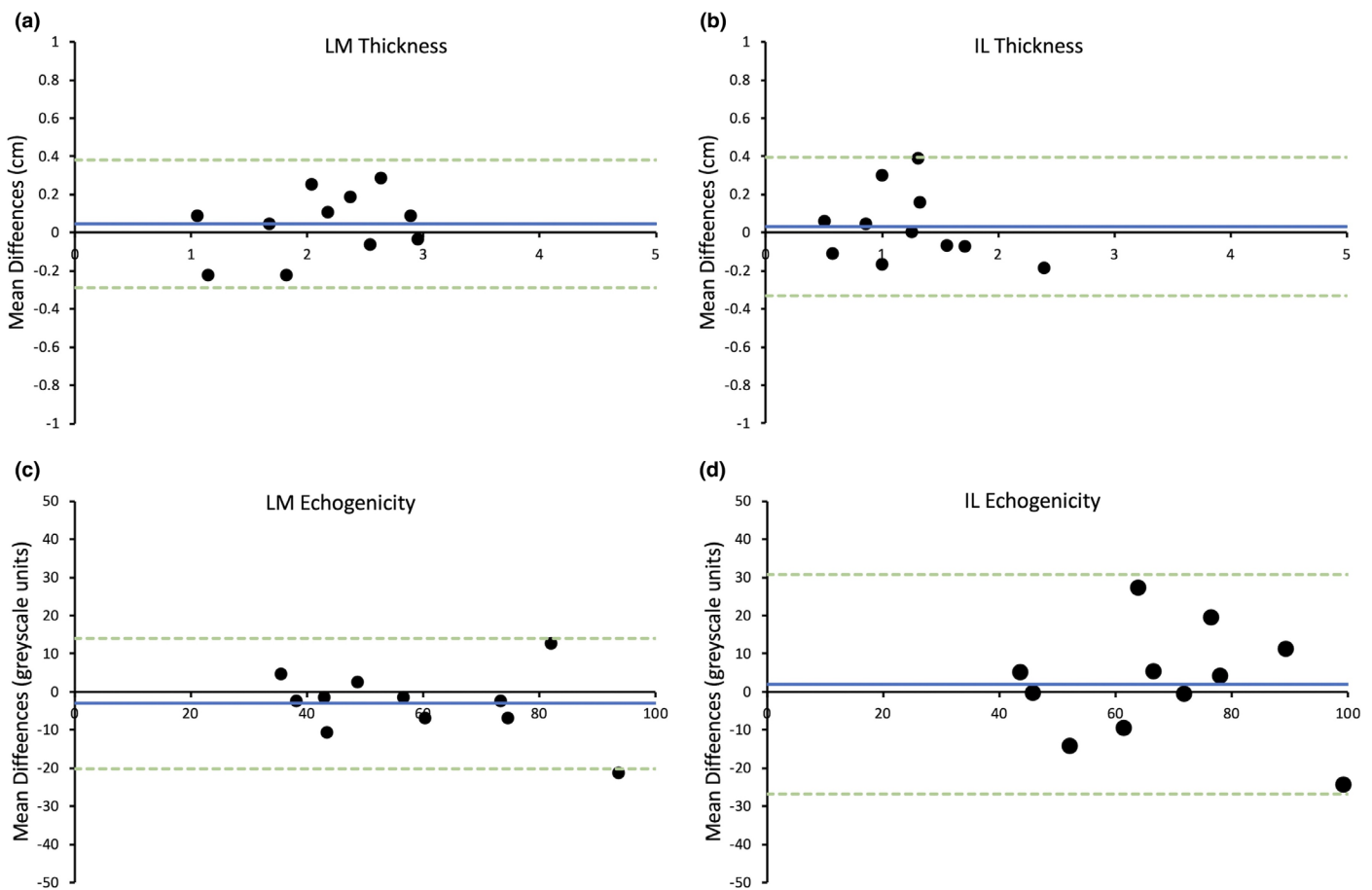


FIGURE 2: Bland–Altman plot displaying limits of agreement between average measurements taken 1 week apart for Assessor A for the thickness of lumbar multifidus (LM) (a) and iliocostalis lumborum (IL) (b) and the echogenicity of LM (c) and IL (d). Blue line = mean differences, green dotted lines = upper and lower limits of agreement (2 standard deviations)

thickness and echogenicity. Moreover, test–retest reliability for IL muscle echogenicity was excellent for both assessors. However, test–retest reliability for IL muscle thickness was poor for Assessor A but excellent for Assessor B. Finally, inter-rater reliability for IL muscle thickness and echogenicity was fair.

Only one previous ultrasound study has investigated the test–retest reliability of LM muscle thickness. Specifically, Cuelar *et al.*³⁸ analysed ultrasound images from older healthy adults and reported excellent test–retest reliability ($ICC_{3,1} = 0.83$) for LM thickness. Likewise, our results support the use of ultrasound imaging as a reliable clinical and research tool for comparison of LM thickness between two separate time points. Furthermore, measures of LM thickness are comparable between assessors with differing amounts of experience using ultrasound. These findings are supported by previous studies that have similarly reported excellent inter-rater reliability.^{2-4,27,39} Importantly, these studies used identical methods for image acquisition and similar training time for assessors to that of the current study.

There was considerable variability in test–retest reliability for IL thickness between assessors. Compared to the LM, the IL has some distinct anatomical features that likely contributed to a reduced level of reliability for thickness measures. For example, the IL varies in its orientation as it descends from the ribs to the lumbar region to attach to the ilium.⁴⁰ As such, the thickness of the IL is different depending on the spinal level at which it is imaged. Another difference is the proximity of the IL to easily identifiable bony landmarks; specifically, the LM sits adjacent to the spinous processes of the lumbar vertebrae passing from the transverse processes up to the spinous processes, whereas the IL is more lateral exhibiting greater variability, making it difficult to reliably locate. Imaging IL requires knowledge of lumbar surface anatomy and experience palpating the lumbar spine. In addition, accurately locating the IL is more challenging in people with higher BMI given that greater amounts of subcutaneous fat increase the required ultrasound capture depth with a resulting decline in image resolution.⁴¹ Higher BMI has also been associated with increases in intra-

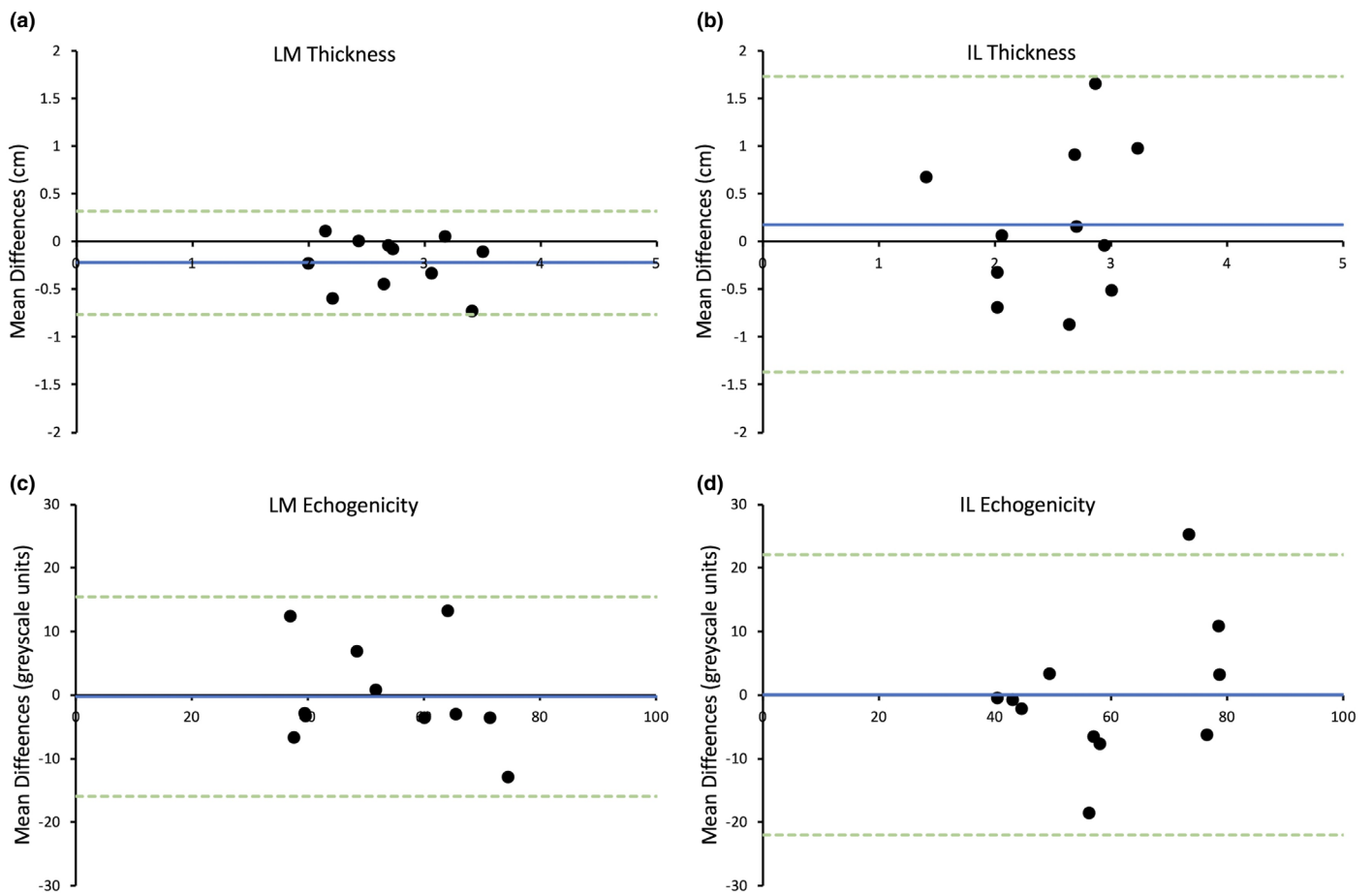


FIGURE 3: Bland–Altman plot displaying difference of agreement between average measurements taken 1 week apart for Assessor B for the thickness of lumbar multifidus (LM) (a) and iliocostalis lumborum (IL) (b) and the echogenicity of LM (c) and IL (d). Blue line = mean differences, green dotted lines = upper and lower limits of agreement (2 standard deviations)

TABLE 3: Inter-rater reliability for LM and IL muscle thickness and echogenicity

Muscle	Thickness				Echogenicity			
	ICC _{2,3}	95% CI	SEM	F	ICC _{2,3}	95% CI	SEM	F
LM	0.79	[0.00, 0.95]	0.15	12.35	0.94	[0.75, 0.98]	0.72	20.63
IL	0.00	[0.00, 0.44]	1.27	0.79	0.39	[0.00, 0.82]	5.25	1.74

Abbreviations: CI, confidence interval; F, F-statistic; ICC, intraclass correlation coefficient; IL, iliocostalis lumborum; LM, lumbar multifidus; SEM, standard error of measurement.

muscular connective tissue,⁴² contributing to greater difficulty identifying muscle and fascial tissue.

Only one previous study has investigated the reliability of echogenicity-related measures of lumbar spine muscles. Specifically, Sarafraz *et al.*⁴³ reported that test–retest reliability of LM echogenicity was excellent (ICC = 0.91–0.94); however, the authors confounded their findings by combining results from both healthy controls and patients with low back pain.

Similarly, high levels of reliability have been observed for quadriceps-related echogenicity measures of healthy and critically ill cohorts (i.e. intra-rater reliability ICC = 0.99; inter-rater reliability ICC = 0.96–0.97, test–retest reliability ICC = 0.88–0.91).^{44,45} Importantly, our findings pertaining to LM-related echogenicity are in agreement with the aforementioned studies. Fat infiltration of the LM has demonstrated a strong association with low back pain in adults.¹⁵ The presence

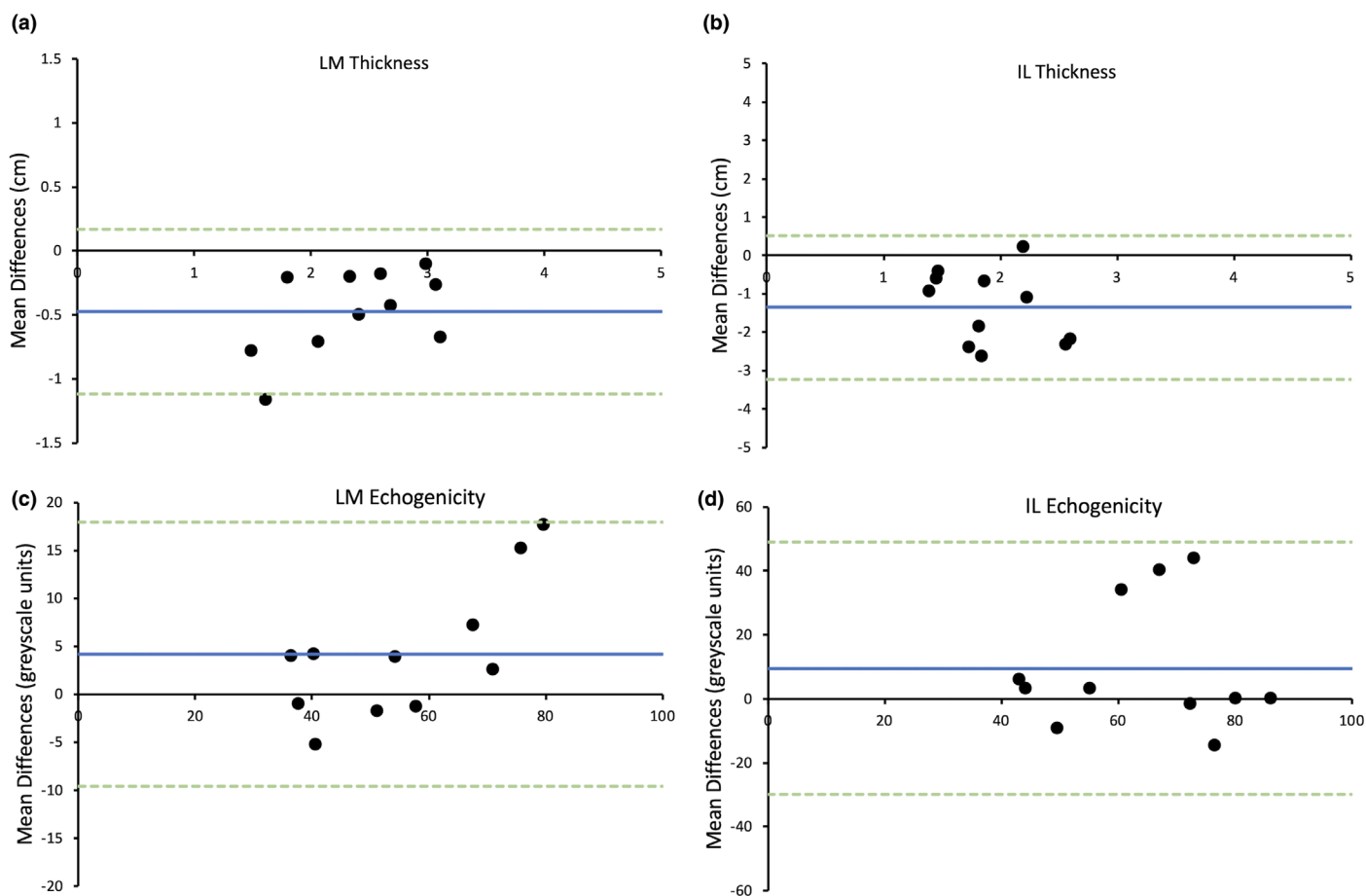


FIGURE 4: Bland–Altman plot displaying difference of agreement between average measurements by the two raters for the thickness of lumbar multifidus (LM) (a) and iliocostalis lumborum (IL) (b) and the echogenicity of LM (c) and IL (d). Blue line = mean differences, green dotted lines = upper and lower limits of agreement (2 standard deviations)

of LM intra-muscular fat is also strongly associated with previous episodes of low back pain,¹⁵ suggesting that fat infiltration of the LM does not spontaneously resolve once the pain has ceased. Whilst Welch, Moran⁴⁶ demonstrated that fat infiltration percentage (relative to the entire muscle area) in the lumbar musculature (i.e. LM and erector spinae) in people with chronic low back pain decreases with resistance training, Berry, Padwal⁴⁷ reported that resistance training is actually limited in its capacity to reduce chronic low back pain-related fatty infiltrates. Differences between studies may be related to the age of the participants. Specifically, the participants in Welch, Moran⁴⁶ were considerably younger (i.e. 39.6 ± 12.4 years) than those in Berry, Padwal⁴⁷ (i.e. 52.8 ± 14.8 years). This is noteworthy because older participants are less likely to demonstrate exercise-related reductions in intra-muscular fat.⁴⁸ Hence, in younger people with chronic low back pain, ultrasound imaging may be a useful tool for clinicians to assess and monitor changes in the composition of LM and IL in response to interventions such as resistance training.

Ultrasound-derived IL echogenicity is only reliable when the same assessor performs the image acquisition and analyses – possibly because of the above-mentioned challenges in locating the IL muscle. The results pertaining to LM and IL echogenicity may enable researchers, rather than referring people for CT (high radiation dose), MRI (expensive) and biopsy (painful), to reliably evaluate the effects of interventions on lumbar muscle composition. This may inform the design of future studies that investigate low back pain intervention and their impact on morphological data and associated variances in functional or disability improvement. However, importantly whilst echogenicity correlates to intra-muscular fat in lower limb muscles,^{20,49} there is no such study to validate this in the lumbar spine.

This study has several limitations. Firstly, this study included a small sample size of participants and assessors, and, as such, this potentially limits the generalisability of our findings. Indeed, to draw strong conclusions, we would need to include multiple assessors and a larger cohort. Secondly, as this study

was conducted on healthy people, we recommend caution when applying the results to low back pain populations. In this respect, future studies should investigate the morphological properties of ultrasound-related outcome measures in a cohort of low back pain patients.

Conclusion

Lumbar multifidus thickness and echogenicity can be assessed reliably by different assessors using ultrasound. IL thickness and echogenicity measurements possess excellent test–retest reliability but poor and fair inter-rater reliability, respectively. Future studies may investigate ultrasound-derived LM and IL morphological features in people with low back pain and the effects of interventions on these muscles.

Acknowledgements

The authors wish to thank the participants for their involvement in the study. We would also like to thank Professor Maria Stokes for her guidance regarding ultrasound imaging acquisition and Mr Samuel Crofts for his advice regarding the ICC statistics within this manuscript.

Authorship Declaration

The authors acknowledge that (i) the authorship listing conforms with the journal's authorship policy, and (ii) that all authors are in agreement with the content of the submitted manuscript.

Funding information

No funding information is provided.

Conflict of Interest

I affirm that I have no financial affiliation (including research funding) or involvement with any commercial organisation that has a direct financial interest in any matter included in this manuscript, except as disclosed in an attachment and cited in the manuscript. Any other conflict of interest (i.e. personal associations or involvement as a director, officer, or expert witness) is also disclosed in an attachment. No conflict of interest has been declared by the authors.

Author Contributions

Joshua Farragher: Conceptualisation (lead); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (lead); Project administration (lead); Writing-original draft (lead); Writing-review & editing (lead). **Adrian Pranata:** Conceptualisation (supporting); Formal analysis (supporting); Methodology (supporting); Supervision (supporting); Validation (equal); Writing-review & editing (equal). **Doa El-Ansary:** Conceptualisation (supporting); Data curation (supporting); Methodology (supporting); Resources (supporting); Supervision (supporting); Validation (supporting); Writing-review & editing (equal). **Selina Parry:** Data curation (supporting);

Methodology (supporting); Software (supporting); Supervision (supporting); Validation (supporting); Writing-review & editing (equal). **Gavin Williams:** Supervision (supporting); Validation (supporting); Writing-review & editing (supporting). **Colin Royse:** Resources (equal); Validation (supporting); Writing-review & editing (supporting). **Alistair Royse:** Resources (equal); Validation (supporting); Writing-review & editing (supporting). **Molly O'Donohue:** Data curation (equal); Formal analysis (equal); Investigation (equal); Writing-review & editing (supporting). **Adam Bryant:** Conceptualisation (supporting); Methodology (supporting); Supervision (lead); Validation (lead); Writing-review & editing (lead).

References

- 1 Cholewicki J, Panjabi MM, Khachatryan A. Stabilizing function of trunk flexor-extensor muscles around a neutral spine posture. *Spine (Phila Pa 1976)* 1997; 22(19): 2207–12.
- 2 Wallwork TL, Stanton WR, Hides JA. Intrarater and interrater reliability of assessment of lumbar multifidus muscle thickness using rehabilitative ultrasound imaging. *J Orthop Sports Phys Ther* 2007; 37: 608–12.
- 3 Kiesel K, Uhl T, Underwood F, Rodd D, Nitz A. Measurement of lumbar multifidus muscle contraction with rehabilitative ultrasound imaging. *Man Ther* 2007; 12: 161–6.
- 4 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(12): 979–88.
- 5 Nabavi N, Mosallanezhad Z, Haghhighatkah HR, Mohseni Bandpeid MA. Reliability of rehabilitative ultrasonography to measure transverse abdominis and multifidus muscle dimensions. *Iran J Radiol* 2014; 11(3): e21008.
- 6 Wilke HJ, Wolf S, Claes LE, Arand M, Wiesend A. Stability increase of the lumbar spine with different muscle groups. *Spine (Phila Pa 1976)* 1995; 20(2): 192–8.
- 7 Freeman MD, Woodham MA, Woodham AW. The role of the lumbar multifidus in chronic low back pain: a review. *Pm r* 2010; 2(2): 142–6. quiz 1 p following 67.
- 8 Hides J, Gilmore C, Stanton W, Bohlscheid E. Multifidus size and symmetry among chronic LBP and healthy asymptomatic subjects. *Man Ther* 2008; 13: 43–9.
- 9 Hides JA, Stokes MJ, Saide M, Jull GA, Cooper DH. Evidence of lumbar multifidus muscle wasting ipsilateral to symptoms in patients with acute/subacute low back pain. *Spine* 1994; 19(2): 165–72.
- 10 Wallwork TL, Stanton WR, Freke M, Hides JA. The effect of chronic low back pain on size and contraction of the lumbar multifidus muscle. *Man Ther* 2009; 14: 496–500.
- 11 Hides JA, Richardson CA, Jull GA. Multifidus muscle recovery is not automatic after resolution of acute, first-episode low back pain. *Spine* 1996; 21(3): 2763–9.
- 12 Steele J, Bruce-Low S, Smith D. A reappraisal of the deconditioning hypothesis in low back pain: review of evidence from a triumvirate of research methods on specific lumbar extensor deconditioning. *Curr Med Res Opin* 2014; 30(5): 865–911.
- 13 Kim HK, Zhang Y. Estimation of lumbar spinal loading and trunk muscle forces during asymmetric lifting tasks: application of

- whole-body musculoskeletal modelling in OpenSim. *Ergonomics* 2017; 60(4): 563–76.
- 14 Bogduk N. A reappraisal of the anatomy of the human lumbar erector spinae. *J Anat* 1980; 131(3): 525–40.
 - 15 Kjaer P, Bendix T, Sorensen JS, Korsholm L, Leboeuf-Yde C. Are MRI-defined fat infiltrations in the multifidus muscles associated with low back pain? *BMC Med* 2007; 5: 2.
 - 16 Wan Q, Lin C, Li X, Zeng W, Ma C. MRI assessment of paraspinal muscles in patients with acute and chronic unilateral low back pain. *Br J Radiol* 2015; 88(1053): 20140546.
 - 17 Hildebrandt M, Fankhauser G, Meichtry A, Luomajoki H. Correlation between lumbar dysfunction and fat infiltration in lumbar multifidus muscles in patients with low back pain. *BMC Musculoskelet Disord* 2017; 18(1): 12.
 - 18 D'hooge R, Cagnie B, Crombez G, Vanderstraeten G, Dolphens M, Danneels L. Increased intramuscular fatty infiltration without differences in lumbar muscle cross-sectional area during remission of unilateral recurrent low back pain. *Man Ther* 2012; 17: 584–8.
 - 19 Ihnatsenka B, Boezaart AP. Ultrasound: basic understanding and learning the language. *Int J Shoulder Surg* 2010; 4(3): 55–62.
 - 20 Harris-Love MO, Avila NA, Adams B, Zhou J, Seamon B, Ismail C, et al. The comparative associations of ultrasound and computed tomography estimates of muscle quality with physical performance and metabolic parameters in older men. *J Clin Med* 2018; 7(10): 340.
 - 21 Wilhelm EN, Rech A, Minozzo F, Radaelli R, Botton CE, Pinto RS. Relationship between quadriceps femoris echo intensity, muscle power, and functional capacity of older men. *Age (Dordr)* 2014; 36(3): 9625.
 - 22 Strasser EM, Draskovits T, Praschak M, Quittan M, Graf A. Association between ultrasound measurements of muscle thickness, pennation angle, echogenicity and skeletal muscle strength in the elderly. *Age (Dordr)* 2013; 35(6): 2377–88.
 - 23 Parry SM, Burtin C, Denehy L, Puthuchery ZA, Bear D. Ultrasound evaluation of quadriceps muscle dysfunction in respiratory disease. *Cardiopulm Phys Ther J* 2019; 30(1): 15–23.
 - 24 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: 347–56.
 - 25 Lariviere C, Gagnon D, De Oliveira E, Jr., Henry SM, Mecheri H, Dumas JP. Ultrasound measures of the lumbar multifidus: effect of task and transducer position on reliability. *PM R* 2013; 5(8): 678–87.
 - 26 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(6): 354–61.
 - 27 Wilson A, Hides JA, Blizzard L, Callisaya M, Cooper A, Srikanth VK, et al. Measuring ultrasound images of abdominal and lumbar multifidus muscles in older adults: a reliability study. *Man Ther* 2016; 23: 114–9.
 - 28 Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hrobjartsson A, et al. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *J Clin Epidemiol* 2011; 64(1): 96–106.
 - 29 Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010; 19(4): 539–49.
 - 30 Stokes M, Rankin G, Newham DJ. Ultrasound imaging of lumbar multifidus muscle: normal reference ranges for measurements and practical guidance on the technique. *Man Ther* 2005; 10(2): 116–26.
 - 31 Farragher JB, Pranata A, Williams G, El-Ansary D, Parry SM, Kasza J, et al. Effects of lumbar extensor muscle strengthening and neuromuscular control retraining on disability in patients with chronic low back pain: a protocol for a randomised controlled trial. *BMJ Open* 2019; 9: e028259.
 - 32 Grani G, D'Alessandri M, Carbotta G, Nesca A, Del Sordo M, Alessandrini S, et al. Grey-scale analysis improves the ultrasonographic evaluation of thyroid nodules. *Medicine (Baltimore)* 2015; 94(27): e1129.
 - 33 Molinari F, Caresio C, Acharya UR, Mookiah MR, Minetto MA. Advances in quantitative muscle ultrasonography using texture analysis of ultrasound images. *Ultrasound Med Biol* 2015; 41(9): 2520–32.
 - 34 Caresio C, Molinari F, Emanuel G, Minetto MA. Muscle echo intensity: reliability and conditioning factors. *Clin Physiol Funct Imaging* 2015; 35(5): 393–403.
 - 35 McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods* 1996; 1(1): 30–46.
 - 36 Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979; 86(2): 420–8.
 - 37 Portney LG, Watkins MP. Foundations of Clinical Research: Applications to Practice, 3rd ed. Upper Saddle River, NJ: Pearson Education, Inc.; 2009.
 - 38 Cuellar WA, Blizzard L, Callisaya ML, Hides JA, Jones G, Ding C, et al. Test-retest reliability of measurements of abdominal and multifidus muscles using ultrasound imaging in adults aged 50–79 years. *Musculoskelet Sci Pract* 2017; 28: 79–84.
 - 39 Sions JM, Velasco TO, Teyhen DS, Hicks GE. Reliability of ultrasound imaging for the assessment of lumbar multifidus thickness in older adults with chronic low back pain. *J Geriatr Phys Ther* 2015; 38(1): 33–9.
 - 40 Macintosh JE, Bogduk N. 1987 Volvo award in basic science. The morphology of the lumbar erector spinae. *Spine (Phila Pa 1976)* 1987; 12(7):658–68.
 - 41 Ng A, Swanevelder J. Resolution in ultrasound imaging. *Contin Educ Anaesthesia Crit Care Pain* 2011; 11(5): 186–92.
 - 42 Langevin HM, Stevens-Tuttle D, Fox JR, Badger GJ, Bouffard NA, Krag MH, et al. Ultrasound evidence of altered lumbar connective tissue structure in human subjects with chronic low back pain. *BMC Musculoskelet Disord* 2009; 10: 151.
 - 43 Sarafraz H, Hadian MR, Yazdi NA, Olyaei G, Bagheri H, Jalaie S, et al. Test-retest reliability of nerve and muscle morphometric characteristics utilizing ultrasound imaging in individuals with unilateral sciatica and controls. *Chiropr Man Therap* 2018; 26: 47.
 - 44 Vieira L, Rocha LPB, Mathur S, Santana L, Melo PF, Silva V, et al. Reliability of skeletal muscle ultrasound in critically ill trauma patients. *Rev Bras Ter Intensiva* 2019; 31(4): 464–73.

- 45 Tomko PM, Muddle TW, Magrini MA, Colquhoun RJ, Luera MJ, Jenkins ND. Reliability and differences in quadriceps femoris muscle morphology using ultrasonography: the effects of body position and rest time. *Ultrasound* 2018; 26(4): 214–21.
- 46 Welch N, Moran K, Antony J, Richter C, Marshall B, Coyle J, *et al.* The effects of a free-weight-based resistance training intervention on pain, squat biomechanics and MRI-defined lumbar fat infiltration and functional cross-sectional area in those with chronic low back. *BMJ Open Sport Exerc Med* 2015; 1(1): e000050.
- 47 Berry DB, Padwal J, Johnson S, Englund EK, Ward SR, Shahidi B. The effect of high-intensity resistance exercise on lumbar musculature in patients with low back pain: a preliminary study. *BMC Musculoskelet Disord* 2019; 20(1): 290.
- 48 Marcus RL, Addison O, Kidde JP, Dibble LE, Lastayo PC. Skeletal muscle fat infiltration: impact of age, inactivity, and exercise. *J Nutr Health Aging* 2010; 14(5): 362–6.
- 49 Young HJ, Jenkins NT, Zhao Q, McCully KK. Measurement of intramuscular fat by muscle echo intensity. *Muscle Nerve* 2015; 52(6): 963–71.
-