The Evaluation of Lumbar Multifidus Muscle Function via Palpation: Reliability and Validity of a New Clinical Test

Jeffrey J. Hebert, DC, PhD Shane L. Koppenhaver, PhD, PT Deydre S. Teyhen, PhD, PT Bruce F. Walker, DC, MPH, DrPH Julie M. Fritz, PhD, PT

PII: S1529-9430(13)01504-0

DOI: 10.1016/j.spinee.2013.08.056

Reference: SPINEE 55553

To appear in: *The Spine Journal*

Received Date: 6 January 2012

Revised Date: 15 April 2013

Accepted Date: 26 August 2013

The Evaluation of Lumbar Multifidus Muscle Function via Palpation: Reliability and Validity of a New Clinical Test

Jeffrey J. Hebert, DC, PhD[1*]

Shane L. Koppenhaver, PhD, PT[2]

Deydre S. Teyhen, PhD, PT[3]

Bruce F. Walker, DC, MPH, DrPH[4]

Julie M. Fritz, PhD, PT[5]

[1] Senior Lecturer, School of Psychology and Exercise Science, Murdoch University, Western Australia

[2] Assistant Professor, U.S. Army-Baylor University Doctoral Program in Physical Therapy, San Antonio, Texas

[3] Associate Professor, U.S. Army-Baylor University Doctoral Program in Physical Therapy, San Antonio, Texas; Commander, U.S. Army Public Health Command Region-South, Fort Sam Houston, Texas

[4] Senior Lecturer, School of Health Professions, Murdoch University, Western Australia

[5] Associate Professor, Department of Physical Therapy, University of Utah and Intermountain Healthcare, Salt Lake City, Utah

This study was supported by the National Institutes of Health, National Center for Complementary and Alternative Medicine (R21 AT004221).

The view(s) expressed herein are those of the author(s) and do not reflect the official policy or position of the U.S. Army Medical Department, the U.S. Army Office of the Surgeon General, the Department of the Army, Department of Defense, or the U.S. Government.

Address for correspondence:

Jeffrey Hebert
Murdoch University
School of Psychology and Exercise Science
Murdoch, Western Australia 6150
Email: J.Hebert@Murdoch.edu.au

The Evaluation of Lumbar Multifidus Muscle Function via Palpation: Reliability and Validity of a New Clinical Test

Abstract

Background Context: The lumbar multifidus muscle provides an important contribution to lumbar spine stability and the restoration of lumbar multifidus function is a frequent goal of rehabilitation. Currently, there are no reliable and valid physical examination procedures available to assess lumbar multifidus function of patients with low back pain.

Purpose: To examine the interrater reliability and concurrent validity of the multifidus lift test to identify lumbar multifidus dysfunction amongst patients with low back pain.

Study Design/Setting: A cross-sectional analysis of reliability and concurrent validity performed in a university outpatient research facility.

Patient Sample: 32 persons aged 18-60 years with current low back pain and a minimum modified Oswestry disability score of 20%. Study participants were excluded if they reported a history of lumbar spine surgery, lumbar radiculopathy, medical red flags, osteoporosis or had recently been treated with spinal manipulation or trunk stabilization exercises.

Outcome Measures: Concurrent measures of lumbar multifidus muscle function at the L4/L5 and L5/S1 levels were obtained with the multifidus lift test (index test) and real-time ultrasound imaging (reference standard).

Methods: The interrater reliability of the multifidus lift test was examined by measuring the level of agreement between two blinded examiners. Concurrent validity of the multifidus lift test was investigated by comparing clinicians' judgements with real-time ultrasound imaging measures of lumbar multifidus function.

Results: Interrater reliability of the multifidus lift test was substantial to excellent (K = 0.75 to 0.81, $p \leq 0.01$) and free from errors of bias and prevalence. When performed at L4/L5 or L5/S1, the multifidus lift test demonstrated evidence of concurrent validity through its relationship with the reference standard results at L4/L5 (r_{bis} = 0.59 to 0.73, $p \leq 0.01$). The multifidus lift test generally failed to demonstrate a relationship with the reference standard results from the L5/S1 level.

Conclusions: Our results provide preliminary evidence supporting the reliability and validity of the MLT to assess lumbar multifidus function at the L4/L5 spinal level. Additional research examining the measurement properties and utility of this test should be undertaken prior to confident implementation with patients.

Key words: Low back pain, diagnosis, reliability, validity, skeletal muscle, spine

Introduction

Low back pain (LBP) is a highly prevalent[1] and costly complaint resulting in substantial socioeconomic burden,[2] with persons experiencing LBP incurring health care expenditures approximately 60% higher than those without LBP pain.[3] When evaluating patients with LBP, traditional diagnostic approaches have focused on the identification of anatomical pain generators. However, this pathoanatomic approach has failed to establish consistent relationships between pathology and symptoms.[4,5]

A different diagnostic approach emphasizes the assessment of function. The lumbar multifidus muscle provides an important contribution to lumbar spine stability.[6-10] Deficits in lumbar multifidus function are associated with LBP[11-13] and the restoration of lumbar multifidus function is a frequent goal of rehabilitation.[14-20] In research, a common method of assessing lumbar multifidus function involves the acquisition of muscle thickness measures using real-time ultrasound imaging and comparing the change in thickness from resting to contracted states.[21,22] However, this technology is expensive and rarely available to clinicians on a routine basis.

Physical examination procedures are a standard aspect of the clinical evaluation and diagnosis of patients with spinal disorders and LBP. Prior to implementation, it is important to understand the psychometric properties of such procedures to elucidate their utility and role in clinical decision-making. However, there is little evidence regarding the validity of many diagnostic tests,[23] and consequently their clinical utility remains poorly defined. Despite evidence suggesting that lumbar multifidus function should be assessed in patients with LBP, there are no physical examination methods with known reliability and validity currently available to clinicians.

Given the high prevalence and socioeconomic burden imposed by LBP and the potential importance of identifying lumbar multifidus muscle dysfunction amongst patients with LBP, it may be valuable for clinicians to access this information about their patients. A simple procedure that does not rely on expensive technology and having acceptable reliability and validity would be most helpful when evaluating lumbar multifidus muscle function. Therefore, the purpose of this study was to examine the interrater reliability and concurrent validity of the multifidus lift test (MLT) to identify lumbar multifidus dysfunction amongst patients with LBP.

Methods

Participants

As part of a larger study examining spinal manipulative therapy,[24] 32 participants between the ages of 18 and 60 years, with current LBP and a minimum modified Oswestry disability score of 20%, were recruited from the University of XXXXXXXXX campus. Potential participants were excluded if they had (1) a history of lumbar spine surgery, (2) signs or symptoms of lumbar radiculopathy, (3) medical “red flags” indicating a potentially serious condition such as cauda equina syndrome, cancer, or infection, (3) osteoporosis, or (4) were recently treated for LBP with spinal manipulative therapy or trunk muscle stabilization exercise. These criteria were chosen to help identify a sample of individuals resembling patients commonly encountered in clinical practice. All participants reviewed and signed consent forms approved by the Institutional Review Board of the University of XXXXXXX (00023996).

Procedures

Once participant eligibility and consent were confirmed, each participant completed self-report measures of their medical history, pain intensity, pain related disability, and fear-avoidance beliefs. LBP intensity was measured using the 0-10 Numeric Pain Rating Scale (NPRS).[25,26] We generated a composite pain intensity score comprised of the average rating between current pain and "best" and "worst" pain intensity in the preceding 24 hours.[27-29] Additionally, the Modified Oswestry Disability Questionnaire[30] was used to estimate LBP related disability. This questionnaire results in scores ranging from 0-100, with higher scores indicating greater disability. Finally, the Fear-Avoidance Beliefs Questionnaire (FABQ)[31] was administered to understand the participant's beliefs regarding their LBP and work and general physical activities.

Reference Standard

All participants underwent testing with a reference standard consisting of lumbar multifidus muscle thickness measures, obtained at the L4/5 and L5/S1 spinal levels, using brightness-mode real-time ultrasound imaging. This technique measures muscle function by examining the relative change in multifidus muscle thickness from rest to a state of submaximal contraction. The change in lumbar multifidus thickness between resting and submaximally contraction states is an indirect assessment of the muscle's automatic function; representing approximately 30% of the maximal voluntary isometric contraction.[21] This measure has identified decrements in LM function among persons with LBP when compared with asymptomatic individuals[32] and following the induction of LBP with hypertonic saline injection.[13] Investigations into the measurement properties of estimating muscle morphology and function with real-time ultrasound have reported good reliability[33,34] and concurrent validity.[21,35]

We used a Sonosite Titan or a Sonosite MicroMaxx imaging system (Sonosite Inc. Bothell, WA) and a 60 mm, 2-5 MHz curvilinear array using methods described by Kiesel et al.[21] Parasagittal images of the lumbar multifidus were obtained at rest and during a submaximal contraction task involving a contralateral arm lift (CAL). Additional details of this procedure have been reported elsewhere.[21,34] Briefly, the participants laid prone on a plinth with a pillow beneath their abdomen to minimize lumbar lordosis. The participants' were positioned with elbows flexed to 90° and shoulders abducted 120° while holding a hand weight normalized to their body mass. Participants weighing less than 68 kg used a .68 kg weight, and those between 68-90 kg or greater than 91 kg used a .91 kg, or 1.36 kg weight respectively. First, a parasagittal image of lumbar multifidus thickness was acquired with the patient relaxed. Next, the thickness measure was repeated while the participants lifted their contralateral arm approximately 5 cm off the table.

To reduce measurement error and increase precision, the mean of 3 measurements in each state was used for analysis.[36] All images were transferred to a personal computer and measured offline at least seven days following acquisition using National Institutes of Health (Bethesda, MD) Image J software (V1.43u). All ultrasound images were acquired and measured by one of the co-authors with five years of ultrasound imaging experience when blinded to participant details and the results of the index test.

Index test

The multifidus lift test (MLT) was performed in a manner similar to the contralateral arm lift during the ultrasound imaging assessment. With the participant relaxed in the prone position, the multifidus muscle was palpated immediately lateral and adjacent to what each examiner believed to be the interspinous space of L4/L5 and L5/S1.[37] We undertook a pragmatic method of identifying the L4/L5 and L5/S1 spinal levels. First, the L4 level was identified by palpating the iliac crests bilaterally and proceeding posteriomedially along the intercristal line to the intersection of the lumbar spine. The level of intersection was considered to be the L3/L4 interspace.[38] Next, the examiner palpated caudally to identify the L4/L5 and L5/S1 interspaces which we used to identify the L4/L5 and L5/S1 spinal levels. These anatomical landmarks are known to vary between individuals,[39] making the correct identification of spinal level by palpation challenging. However, to enhance the external validity of our results, we sought to replicate the conditions consistent with clinical practice environments where more sophisticated imaging options (e.g., ultrasound, fluoroscopy) are not available.

With arms flexed to approximately 120º and elbows flexed to approximately 90º, the patient was instructed to raise their contralateral arm toward the ceiling approximately 5 cm. During the arm lift, the examiner made a qualitative judgment as to whether the participant demonstrated a normal or abnormal lumbar multifidus contraction. This judgment was based upon the degree of contraction as determined by muscle palpation. We operationally defined a normal contraction as one in which a robust and obvious muscle contraction could be palpated during the arm lift. We operationally defined an abnormal contraction as occurring when there was little or no palpable contraction of the muscle during the arm lift. The test result was considered positive when an abnormal muscle contraction was identified and negative when a judgment of normal contraction was made by the examiner. As with the ultrasound assessment, each participant performed this task while holding a hand weight normalized to his or her body mass. Additionally, we were interested in the effects of hand weight use on the index test outcome; therefore, the participants also performed the test without a hand weight.

The order of examiners and weighting condition (arm lift with or without hand weight) were randomly allocated and counterbalanced using simple randomization without replacement. The examiners judged the results of the MLT in a blinded fashion, independent of one another and without knowledge of the reference standard results. Each examiner was a clinician and researcher with more than 10 years of clinical experience and approximately five years of research and ultrasound imaging experience. Aside from achieving consensus on criteria determining a normal and abnormal MLT outcome and strategy for the identification of spinal level, the examiners did not undergo formal training activity.

Data Analyses

Data management and analyses were performed using the Statistical Package for the Social Sciences version 18.0.3 software (SPSS, Chicago, IL). Descriptive statistics, including estimates of central tendency and variability, were calculated to describe the sample of participants and test data.

To examine the interrater reliability of the MLT, we evaluated agreement between two raters by generating raw agreement percentages and Kappa coefficients with 95% confidence intervals. Interrater reliability of the MLT was examined under weighted and unweighted conditions. During the weighted condition, the MLT replicated the ultrasound measures by having the participant hold a small hand weight as described previously. Conversely, the participant did not hold a hand weight during the unweighted condition. Kappa statistics represent the proportion of agreement greater than that expected by chance. While this appears to be a straightforward concept, the interpretation of Kappa coefficients becomes challenging when faced by circumstances with potential to influence the magnitude of the coefficient, namely bias and prevalence.[40] Bias occurs when there is disagreement in the proportion of positive or negative determinations between raters. As bias increases, chance agreement decreases, resulting in inflation of the magnitude of the Kappa coefficient. With large differences in the prevalence of positive and negative determinations, there is increased chance agreement, which lowers the Kappa value. To enhance the interpretation of the Kappa statistics, we calculated indices of prevalence and bias and considered calculating prevalence and bias adjusted Kappa coefficients if high indices were identified.[40] Kappa coefficients are traditionally interpreted as representing excellent agreement above 0.80, substantial agreement between 0.61-0.80, moderate agreement between 0.41-0.60, fair agreement between 0.21-0.40, and slight between 0.00-0.20.[41]

The concurrent validity of the MLT was examined through its relationship with the ultrasound measures of lumbar multifidus function. As with the reliability analysis, participants completed the MLT under weighted and unweighted conditions. We calculated two-tailed biserial correlation coefficients (r_{bis}) between the MLT outcome (positive or negative) and the percentage of thickness change from resting to contracted states ($Thickness_{contracted}$ – $Thickness_{rest}$) / $Thickness_{rest}$). The level of significance for all tests was 0.05. Missing data were not imputed and pairwise deletion was employed.

Results

Participant demographic and clinical characteristics are presented in Table 1. During the ultrasound imaging assessment, we were unable to acquire adequate visualization of the lumbar multifidus with two participants. Additionally, weighted condition MLT data was inadvertently missed on two participants. Therefore, the sample size for the reliability analysis ranged from 30 to 32 participants and the sample size for the validity analysis ranged from 28 to 30 participants.

Results of the interrater reliability analyses are presented in Table 2. There was no appreciable influence of bias or prevalence on the magnitude of the Kappa coefficients. The magnitude of the Kappa results ranged from 0.75 to 0.81, with raw agreement ranging from 86% to 91%. These results indicate substantial to excellent interrater agreement for the MLT. A "worst case scenario" analysis, represented by the lower bound of the confidence intervals, indicated at least moderate agreement for all comparisons. Moreover, as evidenced by the close Kappa coefficient point estimates and substantial overlap of their respective confidence intervals, test agreement did not appear to differ between weighted and unweighted conditions. Therefore, these results represent evidence of substantial to excellent agreement for the MLT test, during the weighted and unweighted conditions.

The overall mean(SD) for the ultrasound measures of percent lumbar multifidus muscle thickness change were 10.36(7.09)% at the L4/L5 level and 5.73(6.04)% at the L5/S1 level. Additional descriptive statistics for the ultrasound measures of lumbar multifidus function, stratified by spinal level, weighting condition and examiner one's MLT results, are presented in table 3.

Analyses of concurrent validity are presented in table 4. The correlation coefficients demonstrated a consistent relationship (r_{bis} = 0.59 to 0.73, $p \leq 0.01$) between the MLT outcome (index test) at L4/L5 and L5/S1 and the ultrasound measures of lumbar multifidus function (reference standard) at L4/L5. Lower levels of LM function were associated with positive MLT test outcomes. With only one exception, the MLT failed to demonstrate a relationship to the ultrasound measures of lumbar multifidus function at L5/S1.

Discussion

In a relatively short period, the paradigm of evidence-based medicine has evolved from a promising concept to the fundamental basis for clinical practice.[42] Evidence should be incorporated into all aspects of patient care, including diagnostic tests. Clinicians use test results to make decisions about diagnosis, therapy selection, and prognosis. Thus, choosing diagnostic tests with acceptable reliability and validity is an important consideration and prerequisite to high-quality patient care.

Although clinicians often assess the lumbar multifidus to make a clinical judgment about its function, little is known about the psychometric properties of these clinical procedures. We sought to examine the interrater reliability and concurrent validity of a clinical test to identify lumbar multifidus muscle dysfunction amongst patients with LBP. Our results demonstrate that the MLT, when performed at the L4/L5 level by the examiners tested in this study, exhibits satisfactory interrater reliability and concurrent validity.

Our results identified several considerations for the clinical application of the MLT. First, loading the contralateral arm with additional weight did not improve the estimates of interrater agreement or validity. Therefore, the use of a hand weight during the MLT is likely to be unnecessary. The original research describing this strategy of automatically activating the lumbar multifidus using a prone contralateral arm lift[21] did identify significant differences in lumbar multifidus activation between the loaded and unloaded conditions. However, our results appear to indicate that such differences in lumbar multifidus activation between loaded and unloaded conditions during a prone contralateral arm lift cannot be appreciated with muscle palpation.

Next, although the MLT was reliable when performed at the L4/L5 and L5/S1 levels, its validity depended on the spinal level assessed. MLT outcomes from either the L4/L5 or L5/S1 levels demonstrated valid estimates of LM function through their relationship with the ultrasound measures from L4/5. However, MLT outcomes did not consistently relate to criterion measures obtained at the L5/S1 level.

Spinal instability is a proposed mechanism of chronic LBP used to justify a range of therapies such as exercise and surgical fusion.[43,44] The lumbar multifidus provides an important contribution to lumbar spine stability[6,8-10] and morphologic change[22,45-48] and diminished function[11-13] of the lumbar multifidus is associated with LBP. Moreover, lumbar multifidus function has been associated with clinical outcome following spinal manipulation[49,50] and predictors of clinical success with spinal stabilization exercise.[51] However, the clinical utility of this knowledge requires the ability of clinicians to implement reliable and valid diagnostic tests to assess lumbar multifidus function. Given the notional importance of the lumbar multifidus and the morphological and structural deficits reportedly associated with LBP, the MLT fills a potentially important need for clinicians to evaluate the function of this muscle.

This study has several strengths and weaknesses that inform the interpretation of our results. We examined the reliability and validity of the MLT in a cohort of individuals who resemble those commonly encountered in clinical practice. Additionally, we employed robust statistical methods to investigate the interrater reliability of the MLT as well as two sources of bias with potential to confound the interpretation of kappa statistics. However, our reliability estimates were derived from repeated measures obtained on the same day. As a result, the stability of these measures over time remains unknown. Due to the continuous scale of measurement of the ultrasound reference standard, we were unable to generate preferred statistics of diagnostic accuracy such as sensitivity, specificity and likelihood ratios. While knowledge of diagnostic accuracy would enhance understanding with respect to the utility of the MLT, and it is possible to dichotomize the results obtained by the ultrasound reference standard, we feel that additional investigation of normal and abnormal ultrasound measures of lumbar multifidus function are required prior to identifying these cut points. A limitation of the study's internal validity relates to our pragmatic approach to identifying the spinal level during the MLT procedure and the inherent error resulting from variation in anatomical landmarks. As a result, we cannot ensure that the spinal level was consistently identified across the MLT and ultrasound comparisons. Finally, the MLT outcomes in this study were obtained by two experienced examiners and may not generalize to individuals with less clinical experience.

These results identify several areas of future research activity. Those being an improved knowledge of ultrasound measures of lumbar multifidus muscle function and specifically the identification of "normal" and "abnormal" cut points to enable the calculation of statistics of diagnostic accuracy for the MLT. Additionally, it should be emphasized that prior to the confident implementation of the MLT, its responsiveness should be examined as should its utility as demonstrated by a positive impact on clinical decision-making and clinical outcomes amongst patients with LBP. Ideally, such standards would be adhered to for all physical examination procedures.

In conclusion, our results provide preliminary evidence supporting the reliability and validity of the MLT to assess LM function at the L4/L5 spinal level. However, as with all physical examination procedures, replication of these results and additional research examining test responsiveness and clinical utility should be undertaken prior to confident implementation of the MLT with patients.

References

1. Walker BF. The prevalence of low back pain: a systematic review of the literature from 1966 to 1998. *J Spinal Disord.* Jun 2000;13(3):205-217.

2. Dagenais S, Caro J, Haldeman S. A systematic review of low back pain cost of illness studies in the United States and internationally. *Spine J.* Jan-Feb 2008;8(1):8-20.

3. Luo X, Pietrobon R, Sun SX, Liu GG, Hey L. Estimates and patterns of direct health care expenditures among individuals with back pain in the United States. *Spine.* Jan 1 2004;29(1):79-86.

4. Abenhaim L, Rossignol M, Gobeille D, Bonvalot Y, Fines P, Scott S. The prognostic consequences in the making of the initial medical diagnosis of work-related back injuries. *Spine.* 1995;20(7):791-795.

5. Deyo RA, Rainville J, Kent DL. What can the history and physical examination tell us about low back pain? *JAMA.* Aug 12 1992;268(6):760-765.

6. Moseley GL, Hodges PW, Gandevia SC. Deep and superficial fibers of the lumbar multifidus muscle are differentially active during voluntary arm movements. *Spine.* 2002;27(2):E29-36.

7. Solomonow M, Zhou BH, Harris M, Lu Y, Baratta RV. The ligamento-muscular stabilizing system of the spine. *Spine.* Dec 1 1998;23(23):2552-2562.

8. Ward SR, Kim CW, Eng CM, et al. Architectural analysis and intraoperative measurements demonstrate the unique design of the multifidus muscle for lumbar spine stability. *J Bone Joint Surg Am.* Jan 2009;91(1):176-185.

9. Wagner H, Anders C, Puta C, et al. Musculoskeletal support of lumbar spine stability. *Pathophysiology.* Dec 2005;12(4):257-265.

10. Wilke HJ, Wolf S, Claes LE. Stability increase of the lumbar spine with different muscle groups: A biomechanical in vitro study. *Spine.* 1995;20:192-198.

11. MacDonald D, Moseley GL, Hodges PW. Why do some patients keep hurting their back? Evidence of ongoing back muscle dysfunction during remission from recurrent back pain. *Pain.* Apr 2009;142(3):183-188.

12. Dickx N, Cagnie B, Achten E, Vandemaele P, Parlevliet T, Danneels L. Changes in lumbar muscle activity because of induced muscle pain evaluated by muscle functional magnetic resonance imaging. *Spine.* Dec 15 2008;33(26):E983-989.

13. Kiesel KB, Uhl T, Underwood FB, Nitz AJ. Rehabilitative ultrasound measurement of select trunk muscle activation during induced pain. *Man Ther.* May 2008;13(2):132-138.

14. Danneels LA, Vanderstraeten GG, Cambier DC, et al. Effects of three different training modalities on the cross sectional area of the lumbar multifidus muscle in patients with chronic low back pain. *Br J Sports Med.* Jun 2001;35(3):186-191.

15. Hides JA, Jull GA, Richardson CA. Long-term effects of specific stabilizing exercises for first-episode low back pain. *Spine.* Jun 1 2001;26(11):E243-248.

16. O'Sullivan PB, Phyty GD, Twomey LT, Allison GT. Evaluation of specific stabilizing exercise in the treatment of chronic low back pain with radiologic diagnosis of spondylolysis or spondylolisthesis. *Spine.* 1997;22(24):2959-2967.

17. Hebert J, Koppenhaver S, Fritz J, Parent E. Clinical Prediction for Success of Interventions for Managing Low Back Pain. *Clinics In Sports Medicine.* 2008;27(3):463-479.

18. Hebert JJ, Marcus RL, Koppenhaver SL, Fritz JM. Postoperative rehabilitation following lumbar discectomy with quantification of trunk muscle morphology and function: a case report and review of the literature. *J Orthop Sports Phys Ther.* Jul 2010;40(7):402-412.

19. Hebert JJ, Koppenhaver SL, Walker BF. Subgrouping Patients With Low Back Pain: A Treatment-Based Approach to Classification. *Sports Health: A Multidisciplinary Approach.* 2011;3(6):534-542.

20. Koppenhaver SL, Fritz JM, Hebert JJ, et al. Association between history and physical examination factors and change in lumbar multifidus muscle thickness after spinal manipulation in patients with low back pain. *J Electromyogr Kinesiol.* Oct 2012;22(5):724-731.

21. Kiesel KB, Uhl TL, Underwood FB, Rodd DW, Nitz AJ. Measurement of lumbar multifidus muscle contraction with rehabilitative ultrasound imaging. *Man Ther.* May 2007;12(2):161-166.

22. 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.* Oct 2009;14(5):496-500.

23. Rubinstein SM, van Tulder M. A best-evidence review of diagnostic procedures for neck and low-back pain. *Best Pract Res Clin Rheumatol.* Jun 2008;22(3):471-482.

24. [Blind for Review]

25. Childs JD, Piva SR, Fritz JM. Responsiveness of the numeric pain rating scale in patients with low back pain. *Spine.* Jun 1 2005;30(11):1331-1334.

26. Jensen MP, Turner JA, Romano JM. What is the maximum number of levels needed in pain intensity measurement? *Pain.* 1994;58(3):387-392.

27. Farrar JT, Berlin JA, Strom BL. Clinically important changes in acute pain outcome measures: a validation study. *J Pain Symptom Manage.* May 2003;25(5):406-411.

28. Jensen MP, Turner JA, Romano JM, Fisher LD. Comparative reliability and validity of chronic pain intensity measures. *Pain.* 1999;83:157-162.

29. Li L, Liu X, Herr K. Postoperative pain intensity assessment: a comparison of four scales in Chinese adults. *Pain medicine (Malden, Mass.* 2007;8(3):223-234.

30. Fritz JM, Irrgang JJ. A Comparison of a Modified Oswestry Disability Questionnaire and the Quebec Back Pain Disability Scale. *Phys Ther.* 2001;81(2):776-788.

31. Waddell G, Newton M, Henderson I, Somerville D, Main CJ. A Fear-Avoidance Beliefs Questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. *Pain.* 1993;52(2):157-168.

32. 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.* Oct 2007;37(10):596-607.

33. 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(23):E848-E856.

34. 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.* Jan 2009;90(1):87-94.

35. Koppenhaver SL, Hebert JJ, Parent EC, Fritz JM. Rehabilitative ultrasound imaging is a valid measure of trunk muscle size and activation during most isometric sub-maximal contractions: a systematic review. *Aust J Physiother.* 2009;55(3):153-169.

36. 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.* Aug 2009;39(8):604-611.

37. Hides J, Scott Q, Jull G, Richardson C. A Clinical Palpation Test to Check the Activation of the Deep Stabilizing Muscles of the Lumbar Spine. *International SportMed Journal.* 2000;1(4).

38. Chakraverty R, Pynsent P, Isaacs K. Which spinal levels are identified by palpation of the iliac crests and the posterior superior iliac spines? *Journal of anatomy.* Feb 2007;210(2):232-236.

39. Snider KT, Kribs JW, Snider EJ, Degenhardt BF, Bukowski A, Johnson JC. Reliability of Tuffier's line as an anatomic landmark. *Spine (Phila Pa 1976).* Mar 15 2008;33(6):E161-165.

40. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *Journal of Clinical Epidemiology.* 1993;46(5):423-429.

41. Landis RJ, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977 1977;33:159-174.

42. Guyatt G. *Users' guides to the medical literature : a manual for evidence-based clinical practice.* 2nd ed. New York: McGraw-Hill Medical; 2008.

43. Frymoyer JW, Selby DK. Segmental instability. Rationale for treatment. *Spine.* 1985;10(3):280-286.

44. Shaughnessy M, Caulfield B. A pilot study to investigate the effect of lumbar stabilisation exercise training on functional ability and quality of life in patients with chronic low back pain. *Int J Rehabil Res.* Dec 2004;27(4):297-301.

45. 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.

46. Mengiardi B, Schmid MR, Boos N, et al. Fat content of lumbar paraspinal muscles in patients with chronic low back pain and in asymptomatic volunteers: quantification with MR spectroscopy. *Radiology.* Sep 2006;240(3):786-792.

47. Barker KL, Shamley DR, Jackson D. Changes in the cross-sectional area of multifidus and psoas in patients with unilateral back pain: the relationship to pain and disability. *Spine.* 2004;29(22):E515-519.

48. Danneels LA, Vanderstraeten GG, Cambier DC, Witvrouw EE, De Cuyper HJ. CT imaging of trunk muscles in chronic low back pain patients and healthy control subjects. *Eur Spine J.* Aug 2000;9(4):266-272.

49. Koppenhaver SL, Fritz JM, Hebert JJ, et al. Association between changes in abdominal and lumbar multifidus muscle thickness and clinical improvement after spinal manipulation. *J Orthop Sports Phys Ther.* Jun 2011;41(6):389-399.

50. Fritz JM, Koppenhaver SL, Kawchuk GN, Teyhen DS, Hebert JJ, Childs JD. Preliminary investigation of the mechanisms underlying the effects of manipulation: exploration of a multivariate model including spinal stiffness, multifidus recruitment, and clinical findings. *Spine (Phila Pa 1976).* Oct 1 2011;36(21):1772-1781.

51. Hebert JJ, Koppenhaver SL, Magel JS, Fritz JM. The relationship of transversus abdominis and lumbar multifidus activation and prognostic factors for clinical success with a stabilization exercise program: a cross-sectional study. *Arch Phys Med Rehabil.* Jan 2010;91(1):78-85.

Table 1. Demographic and clinical characteristics of participants (N = 32)

Characteristic	Value
Age (years)	31.38(12.70)
% female	43.75
BMI (kg/m2)	25.78(5.51)
Oswestry Disability Score (%)	30.31(11.00)
LBP intensity (0-10)	4.42(1.42)
Duration of symptoms (days)	205.00(739.00)†
% with leg pain	12.50
FABQ-PA (0-24)	13.63(4.25)
FABQ-W (0-42)	14.50(9.17)

NOTE: Values are mean (SD) unless otherwise indicated
† Median (interquartile range).
FABQ-PA, fear avoidance beliefs questionnaire physical activity subscale; FABQ-W, fear avoidance beliefs questionnaire work subscale

Table 2. Multifidus lift test interrater reliability

Procedure	Kappa	95% CI	Percent Agreement	Prevalence Index	Bias Index
L4/L5 no weight (N = 32)	0.75*	0.52, 0.97	86	0.06	0.06
L4/L5 with weight (N = 30)	0.79*	0.57, 1.00	90	0.23	0.10
L5/S1 no weight (N = 32)	0.81*	0.62, 1.00	91	0.03	0.09
L5/S1 with weight (N = 30)	0.80*	0.59, 1.00	90	0.10	0.10

* P < 0.001

Table 3. Descriptive statistics of ultrasound measures of % change in lumbar multifidus muscle thickness, stratified by spinal level, weighting condition and MLT result obtained by examiner one.

MLT Procedure	Ultrasound measure of % change in LM thickness at L4/L5*****	Ultrasound measure of % change in LM thickness at L5/S1*****
	Mean(SD)	*Mean(SD)*
L4/L5 MLT (no weight) (N = 30)		
Negative (n = 15)	12.68 (6.39)	6.07 (5.10)
Positive (n = 15)	6.81 (5.56)	4.29 (5.77)
L4/L5 MLT (with weight) (N = 28)		
Negative (n = 16)	12.89 (6.01)	6.79 (5.20)
Positive (n = 12)	5.55 (4.89)	3.02 (5.13)
L5/S1 MLT (no weight) (N = 30)		
Negative (n = 13)	14.23 (5.42)	7.35 (5.66)
Positive (n = 17)	6.84 (5.67)	3.77 (4.92)
L5/S1 MLT (with weight) (N = 28)		
Negative (n = 14)	12.93 (6.05)	6.84 (5.33)
Positive (n = 14)	6.53 (5.64)	3.51 (5.16)

*Performed using weighted condition
MLT, multifidus lift test; LM, lumbar multifidus

Table 4. Multifidus lift test validity

MLT Procedure	***Change in LM thickness at L4/L5***		***Change in LM thickness at L5/S1***	
	r_{bis}	p	r_{bis}	p
Examiner 1				
L4/L5 no weight (N = 30)	***0.59***	*0.010*	0.29	0.201
L4/L5 with weight (N = 28)	**0.71**	0.003	0.44	0.063
L5/S1 no weight (N = 30)	**0.73**	0.002	**0.47**	0.040
L5/S1 with weight (N = 28)	**0.62**	0.008	0.39	0.097
Examiner 2				
L4/L5 no weight (N = 30)	**0.71**	0.002	0.45	0.053
L4/L5 with weight (N = 28)	**0.69**	0.005	0.24	0.341
L5/S1 no weight (N = 30)	**0.69**	0.003	0.44	0.056
L5/S1 with weight (N = 28)	**0.63**	0.009	0.17	0.472

MLT, multifidus lift test; LM, lumbar multifidus; r_{bis}, biserial correlation coefficient