The evaluation of lumbar multifidus muscle function via palpation: Reliability and validity of a new clinical test

ARTICLE in THE SPINE JOURNAL: OFFICIAL JOURNAL OF THE NORTH AMERICAN SPINE SOCIETY · OCTOBER 2013
Impact Factor: 2.43 · DOI: 10.1016/j.spinee.2013.08.056 · Source: PubMed

CITATIONS
6

READS
207

5 AUTHORS, INCLUDING:

Jeffrey J Hebert
Murdoch University
42 PUBLICATIONS 518 CITATIONS
SEE PROFILE

Shane L Koppenhaver
Baylor University
35 PUBLICATIONS 527 CITATIONS
SEE PROFILE

Bruce F Walker
Murdoch University
99 PUBLICATIONS 1,817 CITATIONS
SEE PROFILE

Available from: Jeffrey J Hebert
Retrieved on: 17 November 2015
Clinical Study

The evaluation of lumbar multifidus muscle function via palpation: reliability and validity of a new clinical test

Jeffrey J. Hebert, DC, PhDa,*, Shane L. Koppenhaver, PhD, PTa,b, Deydre S. Teyhen, PhD, PTb,c, Bruce F. Walker, DC, MPH, DrPHd, Julie M. Fritz, PhD, PTe

aSchool of Psychology and Exercise Science, Murdoch University, 90 South St, Perth, Western Australia, Australia
bUS Army-Baylor University Doctoral Program in Physical Therapy, 3599 Scott Rd, San Antonio, TX, USA
cTelemedicine and Advanced Technology Research Center, US Army Medical Research and Material Command, Bldg. 1054 Patchel St, Fort Detrick, MD 21702, USA
dSchool of Health Professions, Murdoch University, 90 South St, Perth, Western Australia, Australia
eDepartment of Physical Therapy, University of Utah and Intermountain Healthcare, 250 S. 1850 E, Salt Lake City, UT, USA

Received 6 January 2012; revised 15 April 2013; accepted 26 August 2013

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 among patients with low back pain.

PURPOSE: To examine the inter-rater reliability and concurrent validity of the multifidus lift test (MLT) to identify lumbar multifidus dysfunction among 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: Thirty-two persons aged 18 to 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 MLT (index test) and real-time ultrasound imaging (reference standard).

METHODS: The inter-rater reliability of the MLT was examined by measuring the level of agreement between two blinded examiners. Concurrent validity of the MLT was investigated by comparing clinicians’ judgments with real-time ultrasound imaging measures of lumbar multifidus function.

RESULTS: Inter-rater reliability of the MLT was substantial to excellent (κ = 0.75 to 0.81, p < .01) and free from errors of bias and prevalence. When performed at L4–L5 or L5–S1, the MLT demonstrated evidence of concurrent validity through its relationship with the reference standard results at L4–L5 (r = 0.59–0.73, p < .01). The MLT generally failed to demonstrate a relationship with the reference standard results from the L5–S1 level.

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 US Army Medical Department, the US Army Office of the Surgeon General, the Department of the Army, Department of Defense, or the US government.

* Corresponding author. School of Psychology and Exercise Science, Murdoch University, Murdoch, Western Australia 6150, Australia. Tel.: (61) 8-9360-2566.
E-mail address: J.Hebert@Murdoch.edu.au (J.J. Hebert)

FDA device/drug status: Not applicable.

Author disclosures: JHH: Grants: National Center for Complementary and Alternative Medicine, National Institutes of Health (F, Paid directly to institution), Murdoch University (B, Paid directly to institution); Other: Sonosite (donation of an ultrasound imaging system for the study). SLK: Nothing to disclose. DST: Grant: National Center for Complementary and Alternative Medicine, National Institutes of Health (F, Paid directly to institution). BFW: Nothing to disclose. JMF: National Center for Complementary and Alternative Medicine, National Institutes of Health (F, Paid directly to institution/employer).

The disclosure key can be found on the Table of Contents and at www.TheSpineJournalOnline.com.
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 before confident implementation with patients. © 2013 Elsevier Inc. All rights reserved.

Keywords: 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. Before 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 among 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 has acceptable reliability and validity would be most helpful when evaluating lumbar multifidus muscle function. Therefore, the purpose of this study was to examine the inter-rater reliability and concurrent validity of the multifidus lift test (MLT) to identify lumbar multifidus dysfunction among 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 Utah 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 Utah (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 to 10 Numeric Pain Rating Scale (NPRS) [25,26]. We generated a composite pain intensity score composed 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 to 100, with higher scores indicating greater disability. Finally, the Fear-Avoidance Beliefs Questionnaire [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–L5 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 lumbar multifidus 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, USA) and a 60-mm, 2- to 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. Additional details of this procedure have been reported elsewhere [21,34]. Briefly, the participants lay 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 0.68-kg weight, and those between 68 and 90 kg or more than 91 kg used a 0.91 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 three measurements in each state was used for analysis [36]. All images were transferred to a personal computer and measured offline at least 7 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 coauthors (with 5 years of research and ultrasound imaging experience). Each examiner was a clinician and researcher with more than 10 years of clinical experience and approximately 5 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.

**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 posteromedially along the intercrystal 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 among 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 (eg, ultrasound, fluoroscopy) are not available.

With arms flexed to approximately 120° and elbows flexed to approximately 90°, the patient was instructed to raise his or her 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 on 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 5 years of research and ultrasound imaging experience. 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, Inc., Chicago, IL, USA). Descriptive statistics, including estimates of central tendency and variability, were calculated to describe the sample of participants and test data.

To examine the inter-rater reliability of the MLT, we evaluated agreement between two raters by generating raw agreement percentages and Kappa coefficients with 95% confidence intervals. Inter-rater 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. Although 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 and 0.80, moderate agreement between 0.41 and 0.60, fair agreement between 0.21 and 0.40, and slight agreement between 0.00 and 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 .05. Missing data were not imputed and pairwise deletion was used.

**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 were 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 inter-rater 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 (standard deviation) 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<=.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 lumbar multifidus 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 inter-rater reliability and concurrent validity of a clinical test to identify lumbar multifidus muscle dysfunction among 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 inter-rater 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 inter-rater 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 lumbar multifidus function through their relationship with the ultrasound measures from L4–L5. 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 [24,49] and predictors of clinical success with spinal stabilization exercise [50]. 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 used robust statistical methods to investigate the inter-rater 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. Because of 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. Although knowledge of diagnostic accuracy would

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

<table>
<thead>
<tr>
<th>Ultrasound measure of % change in LM thickness at L4–L5,*</th>
<th>Ultrasound measure of % change in LM thickness at L5–S1,*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MLT procedure</strong></td>
<td><strong>mean (SD)</strong></td>
</tr>
<tr>
<td>L4–L5 MLT (no weight) (n=30)</td>
<td>Negative (n=15)</td>
</tr>
<tr>
<td></td>
<td>Positive (n=15)</td>
</tr>
<tr>
<td>L4–L5 MLT (with weight) (n=28)</td>
<td>Negative (n=16)</td>
</tr>
<tr>
<td></td>
<td>Positive (n=12)</td>
</tr>
<tr>
<td>L5–S1 MLT (no weight) (n=30)</td>
<td>Negative (n=13)</td>
</tr>
<tr>
<td></td>
<td>Positive (n=17)</td>
</tr>
<tr>
<td>L5–S1 MLT (with weight) (n=28)</td>
<td>Negative (n=14)</td>
</tr>
<tr>
<td></td>
<td>Positive (n=14)</td>
</tr>
</tbody>
</table>

* Performed using weighted condition.

**Table 4**
Multifidus lift test validity

<table>
<thead>
<tr>
<th>Ultrasound measure of % change in LM thickness at L4–L5</th>
<th>Ultrasound measure of % change in LM thickness at L5–S1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MLT procedure</strong></td>
<td><strong>r_{bis}</strong></td>
</tr>
<tr>
<td>Examinee 1</td>
<td></td>
</tr>
<tr>
<td>L4–L5 no weight (n=30)</td>
<td>0.59</td>
</tr>
<tr>
<td>L4–L5 with weight (n=28)</td>
<td>0.71</td>
</tr>
<tr>
<td>L5–S1 no weight (n=30)</td>
<td>0.73</td>
</tr>
<tr>
<td>L5–S1 with weight (n=28)</td>
<td>0.62</td>
</tr>
<tr>
<td>Examinee 2</td>
<td></td>
</tr>
<tr>
<td>L4–L5 no weight (n=30)</td>
<td>0.71</td>
</tr>
<tr>
<td>L4–L5 with weight (n=28)</td>
<td>0.69</td>
</tr>
<tr>
<td>L5–S1 no weight (n=30)</td>
<td>0.69</td>
</tr>
<tr>
<td>L5–S1 with weight (n=28)</td>
<td>0.63</td>
</tr>
</tbody>
</table>

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

Bolded values are statistically significant.
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 before 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 before 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 among 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 lumbar multifidus 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 before confident implementation of the MLT with patients.

References