Reliability of the Goutallier Classification in Quantifying Muscle Fatty Degeneration in the Lumbar Multifidus Using Magnetic Resonance Imaging

Patrick J. Battaglia, DC,[a] Yumi Maeda, DDS, PhD,[b,c] Aaron Welk, DC,[d] Brad Hough, PhD,[e] and Norman Kettner, DC[f]

Abstract

Objective: The purpose of this study was to investigate the reliability of the Goutallier classification system (GCS) for grading muscle fatty degeneration in the lumbar multifidus (LM) using magnetic resonance imaging (MRI) examinations.

Methods: Lumbar spine MRI scans were obtained retrospectively from the radiology department imaging system. Two examiners (a chiropractic diagnostic imaging resident and a board certified chiropractic radiologist with 30 years of experience) independently graded each LM at the L4/5 and L5/S1 intervertebral level. ImageJ pixel analysis software (version 1.47; National Institutes of Health, Bethesda, MD) was used independently by 2 observers to quantify the percent fat of the LM and allow correlation between LM percent fat and GCS grade. Twenty-five subject MRIs were randomly selected. Magnetic resonance imaging scans were included if they were obtained using a 1.5 T imaging system and were excluded if there was evidence of spinal infection, tumor, fracture, or postoperative changes. For all tests, $P < .05$ was defined as significant.

Results: Intraobserver reliability grading LM fat ranged from a weighted κ (κ_w) of 0.71 to 0.93. Mean interobserver reliability grading LM fat was κ_w, 0.76 to κ_w, 0.85. There was a significant ($P < .001$) correlation between LM percent fat and GCS grade. Furthermore, interobserver reliability in determining percent fat was between intraclass correlation coefficient, 0.73 to intraclass correlation coefficient, 0.90.

Conclusions: In this study, the GCS was reliable in grading LM fatty degeneration and correlated positively with a quantified percent fat value. In addition, ImageJ software (National Institutes of Health) was reliable between raters when quantifying LM percent fat. (J Manipulative Physiol Ther 2014;37:190-197)

Keywords: *Magnetic Resonance Imaging; Low-Back Pain; Skeletal Muscle*

[a] Diagnostic Imaging Resident, Department of Radiology, Logan University, Chesterfield, MO.
[b] Research Fellow, Adjunct Professor, Department of Radiology, Logan University, Chesterfield, MO.
[c] Research Fellow, Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, Charlestown, MA.
[d] Diagnostic Imaging Fellow, Department of Radiology, Logan University, Chesterfield, MO.
[e] Chief Information Officer, Department of Information Technology Services, Logan University, Chesterfield, MO.
[f] Chair, Department of Radiology, Logan University, Chesterfield, MO.
Submit requests for reprints to: Patrick J. Battaglia, DC, Logan University, Department of Radiology, 1851 Schoettler Rd. Chesterfield, MO, 63006. (e-mail: *Patrick.battaglia@logan.edu*).
Paper submitted October 15, 2013; in revised form December 13, 2013; accepted December 15, 2013.
0161-4754/$36.00

http://dx.doi.org/10.1016/j.jmpt.2013.12.010

Mechanical low-back pain (LBP) is frequently associated with segmental instability.[1] Management typically includes motor reeducation programs targeting the deep stabilizing muscles of the spine including the lumbar multifidus (LM) musculature, although there have been recent criticisms of the inadequacies of current management of nonspecific LBP, including muscle reeducation strategies, in favor of a multidimensional biopsychosocial perspective.[1,2] Recently, a relationship between atrophy of the LM muscle with subsequent fatty infiltrate/replacement (collectively termed fatty degeneration) and the etiology, remission, and recalcitrant nature of LBP has been emphasized.[3,4]

Several imaging techniques have been used to study fatty degeneration of the LM, including magnetic resonance (MR) spectroscopy, chemical shift magnetic resonance imaging (MRI), multiecho MRI, and computed tomography

(CT).[5-8] Magnetic resonance spectroscopy revealed that the LM in LBP subjects had a significantly higher fat content than asymptomatic controls.[5] Furthermore, greater LM atrophy was present in a study of subjects with LBP and radiculopathy compared with those with only LBP.[9] Recently, a novel MRI approach for quantification of LM fat content was proposed and shown to agree with fat values derived by MR spectroscopy.[7] Currently, there is no established means of grading LM fatty degeneration, and therefore, a reliable and convenient approach would be welcome in routine radiologic and clinical practice and could enhance LBP research methods. A grading classification for fatty degeneration of the rotator cuff muscles using CT imaging was initially proposed by Goutallier et al[10] in 1994. The Goutallier classification system (GCS), in which intramuscular fat of rotator cuff muscles was graded using a semiquantitative scale of 0 to 4, has demonstrated reliability characteristics ranging from moderate to substantial and excellent.[11-13] In addition, GCS grading of fat density for all 4 rotator cuff muscles correlated with the reference standard Hounsfield units on CT arthrography.[13] Magnetic resonance imaging applications of the GCS have been performed for the rotator cuff and gluteal muscles.[14,15]

The adaptation of the GCS scale for semiquantification of LM fat using MRI has yet to be evaluated. Yanik et al[6] applied the GCS to the LM in their study to generate different cohorts that would allow them to study LM fat with chemical shift MRI. However, this study used only one examiner and also did not test for intraobserver reliability when applying the GCS. Furthermore, they also presented no cases of a GCS grade 4.[6] It is thus clear that implementation of a reliable and convenient method in the evaluation of LM fat is desired in both clinical and research settings and will contribute to improved understanding of segmental instability as a mechanism for LBP. Therefore, the primary aim of this study was to examine both intra and interobserver reliability of the GCS when applied to the LM using MRI. The secondary aim was to measure correlation of each observer's grade to a quantified percent fat reference value for a particular LM derived by pixel analysis of axial MRI scans using ImageJ software (version 1.47; National Institutes of Health, Bethesda, MD) for image analysis. The final aim was to assess interobserver reliability when calculating LM percent fat using ImageJ software.

Methods

Selection of Subject Images

After approval from the Logan University Institutional Review Board, a sample of lumbar spine MRI scans was obtained retrospectively through the department of radiology picture archiving and communication system. The picture archiving and communication system data set consists of images obtained at both hospitals and imaging centers. In an effort to homogenize the sample, MRI examinations were only included if they were performed using a 1.5T GE system (GE Healthcare, Milwaukee, WI) and were excluded if there was evidence of spinal infection, tumor, fracture, or postoperative changes. Also excluded were scans in which the slice plane did not extend through the intervertebral disc. Degenerative disc disease was not an exclusion criterion. Examiners were blinded to all demographic and clinical subject information. Twenty-five MRI examinations of different subjects were selected. Images reflecting all 5 GCS grades (0-4) were chosen (Figure 1). Analysis was performed using T1-weighted (T1W) axial images from the midportion of the L4/5 and L5/S1 intervertebral discs. These levels were chosen because they have been shown to contain the highest LM volume and relative percent fat in healthy controls.[16] Scans were deidentified, assigned a subject number, and exported to a separate file. If a discrepancy arose in which either the image was not acquired through the midplane of the intervertebral disc or if the image plane was acquired through the disc plane of a lumbar scoliosis distorting the anatomical plane, the slice above or below yielding a more symmetrical image plane was selected.

Semiquantification of LM Fat

Two examiners (one first-year chiropractic diagnostic imaging resident and one board certified chiropractic radiologist with 30 years of experience interpreting lumbar spine MRIs; both examiners are authors of this article) independently applied the GCS to the LM. The examiners were blinded to the findings of other raters during the study. The GCS is defined as: grade 0, normal muscle; grade 1, fatty streaks within the muscle; grade 2, fat less than muscle; grade 3, fat and muscle equal; and grade 4, fat greater than muscle. Examiners separately graded the right and left LM for fat content at both the L4/5 and L5/S1 levels on T1W MR axial images. Examiners were given a sheet of paper with the GCS written on it for their grading sessions. No other preparatory training (eg, consensus grading of other lumbar MRIs) was performed. Magnetic resonance images were assessed using a diagnostic grade work station monitor in an optimally lit and quiet room. Examiners were instructed to consider only the LM and not the adjacent erector spinae muscles. Examiners were also instructed to exclude the fat containing fascial plane (cleavage plane) between these 2 muscles in their GCS assessment. The GCS assessment was repeated for each examiner with a 7-day interval between sessions to assess both intra and interobserver reliability. Raters were blinded to their previous findings before their repeat session.

Quantification of LM Fat

To provide a reference standard for correlation, quantification of LM fat content was performed with