RELIABILITY OF THE GOUTALLIER CLASSIFICATION SYSTEM IN THE EVALUATION OF THE MULTIFIDUS MUSCLE

CONCORDÂNCIA DA CLASSIFICAÇÃO DE GOUTALLIER NA AVALIAÇÃO DO MÚSCULO MULTÍFIDO

FIABILIDAD DEL SISTEMA DE CLASIFICACIÓN DE GOUTALLIER EN LA EVALUACIÓN DEL MÚSCULO MULTÍFIDO

Ramon Oliveira Soares[1], Nelson Astur[1], Fabio Chaud de Paula[1], Paulo Simões Forte[2], Guilherme Alves de Melo[2], Marcelo Wajchenberg[1], Pedro Pohl[1], Délio Eulálio Martins[2]

1. Hospital Israelita Albert Einstein, São Paulo, SP, Brazil.
2. Universidade Anhembi Morumbi, São Paulo, SP, Brazil.

ABSTRACT

Introduction: The paravertebral musculature is essential for the biomechanics and stability of the spine, and its involvement in the pathophysiology of spinal diseases has been demonstrated. Qualitative evaluation of muscle degeneration is usually performed by analyzing the fat infiltration rate proposed by the Goutallier classification system. Objective: The objective of this study is to analyze the intra- and interobserver agreement of the Goutallier Classification for the evaluation of fatty degeneration of the multifidus muscle, using magnetic resonance imaging exams. Methods: The study included 68 patients, all diagnosed with symptomatic disc hernia and indicated for surgery. Preoperative magnetic resonance images were used for the analyses. The images were initially evaluated by two orthopedists and two medical students, and then re-evaluated after two weeks. Intra- and inter-observer reliability analysis was performed using the Fleiss Kappa test and the Landis and Koch criteria. All the analyses were performed using the R statistical environment (R Development Core Team, version 3.3.1, 2016) and the significance level was set at 5%. Results: The percentages of intra- and inter-observer agreement were 86.76% and 61.03%, respectively. The intraobserver agreement was near perfect and the interobserver agreement was moderate. Conclusion: The Goutallier Classification System showed moderate interobserver and intraobserver agreement, being a relevant tool for the evaluation of paravertebral musculature fat replacement. ***Level of evidence II; Prospective study for diagnostic purposes.***

Keywords: Paraspinal Muscles; Low Back Pain; Muscular Atrophy.

RESUMO

Introdução: A musculatura paravertebral é essencial para a biomecânica e estabilidade da coluna e tem sido demonstrado seu envolvimento na fisiopatologia das doenças da coluna vertebral. A avaliação qualitativa da degeneração muscular é usualmente feita pela análise da taxa de infiltração de gordura proposta pelo Sistema de Classificação de Goutallier. Objetivo: O objetivo deste trabalho é analisar a concordância intra e interobservador da Classificação de Goutallier para avaliação da degeneração gordurosa do músculo multífido por meio de exames de ressonância magnética. Métodos: Fizeram parte do estudo 68 pacientes, todos com diagnóstico de hérnia discal sintomática e com indicação cirúrgica. As imagens de ressonância magnética colhidas no pré-operatório foram analisadas por dois ortopedistas e dois estudantes de medicina e foram reavaliadas duas semanas depois. Foi realizada análise de confiabilidade intra e interobservador por meio do teste Kappa de Fleiss e pelos critérios de Landis e Koch. Todas as análises foram realizadas com o auxílio do ambiente estatístico R (R Development Core Team, versão 3.3.1, 2016), e o nível de significância foi fixado em 5%. Resultados: As porcentagens de concordância intra e interobservadores foram, respectivamente, 86,76% e 61,03%. A concordância intraobservador foi quase perfeita e moderada interobservadores. Conclusões: O Sistema de Classificação de Goutallier demonstrou concordância moderada interobservador e intraobservador, sendo uma ferramenta relevante na avaliação da substituição gordurosa da musculatura paravertebral. ***Nível de evidência II; Estudo prospectivo para fins diagnósticos.***

Descritores: Músculos Paraespinais; Dor Lombar; Atrofia Muscular.

RESUMEN

Introducción: La musculatura paravertebral es fundamental para la biomecánica y la estabilidad de la columna y se ha demostrado su intervención en la fisiopatología de las enfermedades de la columna. La evaluación cualitativa de la degeneración muscular se suele realizar analizando la tasa de infiltración de grasa propuesta por el sistema de clasificación de Goutallier. Objetivo: El objetivo de este estudio es analizar la concordancia intra e interobservador de la Clasificación de Goutallier para evaluar la degeneración grasa del músculo multífido mediante resonancia magnética. Métodos: Formaron parte del estudio 68 pacientes, todos con diagnóstico de hernia discal sintomática y con indicación quirúrgica. Se utilizaron imágenes de resonancia magnética preoperatorias para el análisis. Las imágenes fueron evaluadas

*inicialmente por 02 ortopedistas y 02 estudiantes de medicina y, después de dos semanas, reevaluadas. El análisis de fiabilidad intra e interobservador se realizó mediante la prueba Kappa de Fleiss y los criterios de Landis y Koch. Todos los análisis se realizaron utilizando el entorno estadístico R (R Development Core Team, 2016), versión 3.3.1, y el nivel de significancia se estableció en 5%. Resultados: Los porcentajes de concordancia intra e interobservador fueron, respectivamente, 86,76% y 61,03%. La concordancia intraobservador fue casi perfecta y la concordancia interobservador fue moderada. Conclusión: el Sistema de Clasificación de Goutallier demostró una moderada concordancia interobservador e intraobservador, siendo una herramienta relevante en la evaluación del reemplazo graso de la musculatura paravertebral. **Nivel de evidencia II; Estudio prospectivo con fines diagnósticos.***

Descriptores: Músculos Paraespinales; Dolor de la Región Lumbar; Atrofia Muscular.

INTRODUCTION

The paravertebral musculature, especially the multifidus muscle, plays a fundamental role in lumbar spine stabilization and is important for the maintenance of vertebral spine biomechanics.[1,2] It is well established that degeneration of the multifidus muscle is often associated with symptoms of lumbar pain, leading to poorer functional results and higher health costs.[3] Its possible involvement as one of the mediators in the pathophysiology of degenerative spine diseases, chronic lumbar pain, and idiopathic scoliosis has been demonstrated.[4–6]

One of the most frequently discussed methods for qualitative evaluation of muscle degeneration is the rate of fat infiltration.[7] Among the different imaging methods tested, magnetic resonance imaging was the one that showed the best degree of agreement with the intramuscular adipose tissue, detected by histological exam.[8] The systems developed, although useful and accurate, are limited by the lengthy analysis process and often, the need to use computer software, which limits the routine muscular evaluation in the clinical environment.

The Goutallier classification system, reported in 1994, was the first system described to classify muscles by the amount of fat within them.[9] Commonly used by experts in shoulder diseases, the system grades intramuscular fat in the rotator cuff muscles using a semi-quantitative scale of 0 to 4, demonstrating reliability ratios ranging from moderate to optimal.[10] Recently, the Goutallier classification has been applied to the analysis of the gluteus muscles in magnetic resonance imaging, with excellent intra- and interobserver agreement.[11]

Adaptation of the system for qualitative analysis of degeneration of the multifidus muscle has been described by some researchers, presenting a correlation with the rate of fat infiltration.[8,12]1 in 20 patients (patient group 1 However, the short interval between analyses, the lack of randomization of the images for the observers' evaluations, the small sample size, and the heterogeneity of the sample are all factors that can influence the results obtained. The implementation of a reliable and convenient method for qualitative evaluation of fat infiltration in the multifidus muscle is desirable in both clinical and research settings. The objective of this study is to analyze the intra- and interobserver agreement of the Goutallier Classification for the evaluation of fatty degeneration of the multifidus muscle by means of magnetic resonance exams.

METHODOLOGY

This is a diagnostic study conducted in a private tertiary hospital, after approval of the protocol by the Research Ethics Committee (CAAE 27043519.0.0000.5492). All patients agreed to participate in this research by signing an Informed Consent Form (ICF). The inclusion criteria were as follows: patients between 18 and 65 years of age; diagnosis of lumbar disc hernia; and surgical indication for unilateral single-level decompression. Cases of reoperation, history of infectious disease, oncologic disease or fracture of the vertebral column were excluded, as were those with incomplete medical records.

T2-weighted magnetic resonance images collected in the preoperative period, of the midplane of the intervertebral disk level indicated for surgery, were selected. All the MRI scans were performed between January 2017 and March 2019, extracted from the same digital platform (Carestream Health, Inc. 2020). The images were classified by four examiners, independently of each other: two orthopedists and two students in their final year of the medical course. All the observers were shown the classification system once the research project had been defined for the study. Although familiar with the classification, they did not use it routinely in their clinical practice and had received no previous training on how to avoid possible bias regarding the homogenization of the results. The cases were initially evaluated by all the participants, and the examiners did not have access to the findings of the other observers during the study. After two weeks, the analyses were repeated by the same examiners, but in a different order of cases than in the first evaluation. All the observers were instructed to disregard, for the purposes of their analyses, both the muscles adjacent to the multifidus and the fat contained in the fascial plane.

The Goutallier classification system (GCS) consists of a semi-quantitative scale of 0 to 4, defined as: grade 0, without infiltration of intramuscular fat; grade 1, some fatty streaks present; grade 2, evident fat, but proportionally less than the muscle tissue; grade 3, amounts of fat equal to the amount of muscle; grade 4, more fat than muscle. (Figure 1)

In order to verify the intra- and interobserver agreement of the Goutallier Classification, we opted to use the Fleiss Kappa test,[13] a generalization of the Kappa test, which is appropriate for situations where several observers are evaluating the same sample on a scale with several categories, as is the case with the Goutallier Classification, which can assume values from 0 to 4. The Kappa coefficient of agreement has values ranging from +1 (perfect agreement) to 0 (agreement equivalent to chance) to –1 (complete disagreement).

The analysis of the results, after calculating the Kappa, was interpreted according to the criteria proposed by Landis and Koch: < 0.00 poor; 0.00 – 0.20 slight; 0.21 – 0.40 fair; 0.41 – 0.60 moderate; 0.61 – 0.80 substantial; 0.81 – 1.00 almost perfect. All the analyses were conducted with the help of the R (R Development Core Team, 2016) statistical environment, version 3.3.1, and the significance level was set at 5%.
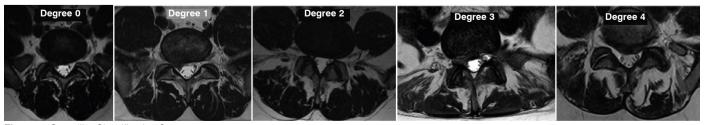

Figure 1. Goutallier Classification System.

RESULTS

In all, 68 patients were included in the study, of which 36 were male and 32 female. The mean age of the participants was 46 ± 14 years, and 53.1% were men. 544 images were included in the analysis, and according to the GCS: 33 were grade 0; 152 grade 1; 197 grade 2; 75 grade 3; and 19 grade 4. (Figure 2)

The interobserver reproducibility study (Table 1) was performed by comparing the first and second readings of each observer, in pairs, in order to cover all possible combinations, analyzed by the Fleiss kappa test. In this analysis, the mean percentage of agreement was 61.03% (standard deviation of 5.16%).

The intra-observer reproducibility study (Table 2), performed using the Fleiss kappa test, showed a mean percentage of agreement of 86.76% (standard deviation of 13.32%).

DISCUSSION

The lumbar paravertebral muscles provide dynamic movement and stability to the spine. There are growing research interests in its normal physiology and its pathological states, and several studies have been carried out to analyze the evaluation parameters.[14,15]and (2 This study aimed to evaluate the reproducibility of the Goutallier Classification in the analysis of the rate of fat infiltration of the multifidus muscle through magnetic resonance imaging, and presented
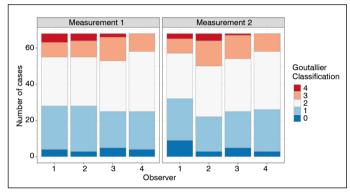

Figure 2. Frequency distribution of the Goutallier classification attributed by observers, by measurement.

Table 1. Results of the interobserver agreement analysis by the Goutallier classification.

Measurement	Observer	Agreement	Fleiss Kappa	Level of reliability
Measurement 1	Observer 1x2	64.71%	0.49	Moderate
	Observer 1x3	64.71%	0.50	Moderate
	Observer 1x4	52.94%	0.30	Fair
	Observer 2x3	66.18%	0.51	Moderate
	Observer 2x4	57.35%	0.36	Fair
	Observer 3x4	60.29%	0.41	Moderate
Measurement 2	Observer 1x2	63.24%	0.49	Moderate
	Observer 1x3	58.82%	0.42	Moderate
	Observer 1x4	54.41%	0.33	Fair
	Observer 2x3	60.29%	0.43	Moderate
	Observer 2x4	58.82%	0.39	Fair
	Observer 3x4	60.29%	0.40	Fair

Table 2. Results of the intraobserver agreement analysis of the Goutallier classification.

Observer	Agreement	Fleiss Kappa	Level of reliability
1	69.12%	0.56	Moderate
2	83.82%	0.77	Substantial
3	97.06%	0.96	Almost perfect
4	97.06%	0.95	Almost perfect

moderate inter- and almost perfect intraobserver agreement. The data in the literature corroborate this relationship. Battaglia et al.,[8] analyzing the multifidus muscle, report that the Goutallier classification is correlated with the rate of fat infiltration, with a correlation coefficient of 0.77-0.89, and inter- and intraobserver reliability of more than 0.8. Yanik et al.,[12]1 in 20 patients (patient group 1 applied the GCS to classify the multifidus muscle in their study to generate different cohorts that would allow them to study the relationship between the fat infiltration rate and chemical displacement, in a magnetic nuclear resonance spectrum record.

Previous studies have classified fatty degeneration using the "mild", "moderate" and "severe" scale, reporting moderate to substantial intra- and interobserver reliability.[16,17] Computed tomography has also been described to evaluate the amount of fat deposits in the paravertebral muscles, through histograms and a semi-quantitative visual grading method. Kalichman et al.,[18] used the mild, moderate and severe scale to classify the multifidus muscle in axial tomography, and reported good intraobserver and moderate to good interobserver reliability. We believe that the adaptation of the GCS for use in the multifidus muscle offers some advantages over the method described previously: the GCS is semiquantitative and provides a numerical scale for fat content, aiding in the interpretation; the magnetic resonance imaging showed an excellent degree of agreement with the intramuscular adipose tissue detected by the histological examination; it is the preferred test for diagnostic investigation of acute and chronic back pain. In addition, when used solely to evaluate the paravertebral muscles, magnetic resonance is less invasive than tomography because it does not use radiation.

Another method described is the evaluation of the cross-section area of the muscle. This is obtained by manually tracing the paravertebral muscle using a computer software program, and then calculating the high intensity area rate within the muscle as a ratio of fat infiltration, using a pseudo-staining technique or histographic analysis.[4,17] This process decreases the practicality of the method, preventing its daily use in the clinical environment. In addition, the cross-section area of the muscle may not undergo significant changes due to fat infiltration in the muscular bundle.[16]

Fatty degeneration has been described in different spine pathologies. Kang et al.,[19] observed higher proportions of fat deposits in the paravertebral muscles in patients with degenerative lumbar kyphosis compared with the control group. Shafaq et al. demonstrated that muscle degeneration was more common on the concave side in patients with degenerative lumbar scoliosis.[20] Stevens et al., conducted a systematic review to evaluate the differences in the morphology of the multifidus muscle between people with lumbar disc hernia and a control group. The involvement of the paraspinal muscles may be a possible mediator in the pathophysiology of disc hernia, with the presence of ipsilateral multifidus muscle changes in people with unilateral lumbar disc hernia.[6]

Although some studies still question whether muscle degeneration is the cause or the result of pathological processes of the spine,[21] its clinical implications have been established. A series of studies found greater fat infiltrate of the multifidus muscle in patients with chronic back pain compared to asymptomatic controls.[16,22,23] The degree of involvement of the paraspinal muscles, especially the multifidus muscle, may be a possible mediator not only in the physiopathology, but also a factor influencing the course of pain and the functional results after the treatment of lumbar pathologies – whether non-surgical or surgical. Therefore, evaluation of the morphology of the multifidus muscle may be an important factor for treatment optimization.

The present study has some limitations. It was conducted in a single center, which could have implications for the extrapolation and reproducibility of the results. However, performing MRI scans at the same imaging center ensures the same pixel intensity profile and acquisition parameters. This, in addition to the fact that all the participants had a diagnosis of symptomatic disc hernia, resulted in a more homogeneous sample. The report of the adapted Goutallier classification and degenerative disc

diseases are valid and reproducible as objects of research, with excellent correlation.[7]

Another important factor is the order in which MRI examinations were presented to the observers, i.e., sequentially. This may have influenced the analyzers during the repeatability session, which may have compromised intraobserver reliability. However, there was a two-week interval between sessions, which minimized the effect of this limitation. Future studies with larger samples may be necessary to validate our results.

CONCLUSION

The Goutallier Classification System showed moderate interobserver and almost perfect intraobserver agreement in the evaluation of the fat infiltration rate of the multifidus muscle.

All authors declare no potential conflict of interest related to this article.

CONTRIBUTIONS OF THE AUTHORS: Each author made significant individual contributions to this manuscript. ROS, NA, DEM: intellectual concept and writing the article; FCP: acquisition and analysis of the data and writing the article; PSF, GAM: acquisition of data and writing the article; MW, PP: review of the article and interpretation of the data.

REFERENCES

1. Altinkaya N, Cekinmez M. Lumbar multifidus muscle changes in unilateral lumbar disc herniation using magnetic resonance imaging. Skeletal Radiol. 2016;45(1):73-7. doi:10.1007/s00256-015-2252-z.
2. Cholewicki J, Panjabi MM, Khachatryan A. Stabilizing Function of Trunk Flexor-Extensor Muscles Around a Neutral Spine Posture. Spine. 1997;22(19):2207-12. doi:10.1097/00007632-199710010-00003.
3. 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. doi:10.1016/j.pmrj.2009.11.006.
4. Lee H Il, Song J, Lee HS, Kang JY, Kim M, Ryu JS. Association between Cross-sectional Areas of Lumbar Muscles on Magnetic Resonance Imaging and Chronicity of Low Back Pain. Ann Rehabil Med. 2011;35(6):852-9. doi:10.5535/arm.2011.35.6.852.
5. Wajchenberg M, Astur N, Fernandes EA, Paredes – Gameiro EJ, Luciano RP, Scmidt B, et al. Assessment of fatty infiltration of the multifidus muscle in patients with adolescent idiopathic scoliosis through evaluation by magnetic resonance imaging compared with histological analysis: A diagnostic accuracy study. J Pediatr Orthop Part B. 2019;28(4):362-7. doi:10.1097/BPB.0000000000000578.
6. Stevens S, Agten A, Timmermans A, Vandenabeele F. Unilateral changes of the multifidus in persons with lumbar disc herniation : a systematic review and meta-analysis. Spine J. 2020;20(10):1573-85. doi:10.1016/j.spinee.2020.04.007.
7. Tamai K, Chen J, Stone M, Arakelyan A, Paholpak P, Nakamura H, et al. The evaluation of lumbar paraspinal muscle quantity and quality using the Goutallier classification and lumbar indentation value. Eur Spine J. 2018;27(5):1005-12. doi:10.1007/s00586-018-5485-4.
8. Battaglia PJ, Maeda Y, Welk A, Hough B, Kettner N. Reliability of the Goutallier Classification in Quantifying Muscle Fatty Degeneration in the Lumbar Multifidus Using Magnetic Resonance Imaging. J Manipulative Physiol Ther. 2014;37(3):190-7. doi:10.1016/j.jmpt.2013.12.010.
9. Goutallier D, Postel JM, Bernageau J, Lavau L, Voisin MC. Fatty muscle degeneration in cuff ruptures. Pre- and postoperative evaluation by CT scan. Clin Orthop Relat Res. 1994;(304):78-83. http://www.ncbi.nlm.nih.gov/pubmed/8020238.
10. Slabaugh MA, Friel NA, Karas V, Romeo AA, Verma NN, Cole BJ. Interobserver and Intraobserver Reliability of the Goutallier Classification Using Magnetic Resonance Imaging. Am J Sports Med. 2012;40(8):1728-34. doi:10.1177/0363546512452714.
11. Engelken F, Wassilew GI, Köhlitz T, Brockhaus S, Hamm B, Perka C, et al. Assessment of Fatty Degeneration of the Gluteal Muscles in Patients With THA Using MRI: Reliability and Accuracy of the Goutallier and Quartile Classification Systems. J Arthroplasty. 2014;29(1):149-53. doi:10.1016/j.arth.2013.04.045.
12. Yanik B, Keyik B, Conkbayir I. Fatty degeneration of multifidus muscle in patients with chronic low back pain and in asymptomatic volunteers: Quantification with chemical shift magnetic resonance imaging. Skeletal Radiol. 2013;42(6):771-8. doi:10.1007/s00256-012-1545-8.
13. Fleiss JL, Levin B, Paik MC. Statistical Methods for Rates and Proportions. John Wiley & Sons, Inc.; 2003.
14. Kulig K, Scheid AR, Beauregard R, Popovich JMJ, Beneck GJ, Colletti PM. Multifidus morphology in persons scheduled for single-level lumbar microdiscectomy: qualitative and quantitative assessment with anatomical correlates. Am J Phys Med Rehabil. 2009;88(5):355-61. doi:10.1097/phm.0b013e31819c506d.
15. Takayama K, Kita T, Nakamura H, Kanematsu F, Yasunami T, Sakanaka H, et al. New predictive index for lumbar paraspinal muscle degeneration associated with aging. Spine. 2016;41(2):E84-90. doi:10.1097/BRS.0000000000001154.
16. 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. doi:10.1186/1741-7015-5-2.
17. Kader DF, Wardlaw D, Smith FW. Correlation Between the MRI Changes in the Lumbar Multifidus Muscles and Leg Pain. Clin Radiol. 2000;55(2):145-9. doi:10.1053/crad.1999.0340.
18. Kalichman L, Klindukhov A, Li L, Linov L. Indices of Paraspinal Muscles Degeneration: Reliabilitu and association with facet joint osteoarthritis: Feasibility study. Clin Spine Surg. 2016;29(9):465-70. doi:10.1097/BSD.0b013e31828be943.
19. Kang CH, Shin MJ, Kim SM, Lee SH, Lee C-S. MRI of paraspinal muscles in lumbar degenerative kyphosis patients and control patients with chronic low back pain. Clin Radiol. 2007;62(5):479-86. doi:10.1016/j.crad.2006.12.002.
20. Shafaq N, Suzuki A, Matsumura A, Terai H, Toyoda H, Yasuda H, et al. Asymmetric Degeneration of Paravertebral Muscles in Patients With Degenerative Lumbar Scoliosis. Spine. 2012;37(16):1398-406. doi:10.1097/BRS.0b013e31824c767e.
21. Mannil M, Burgstaller JM, Thanabalasingam A, Winklhofer S, Betz M, Held U, et al. Texture analysis of paraspinal musculature in MRI of the lumbar spine: analysis of the lumbar stenosis outcome study (LSOS) data. Skeletal Radiol. 2018;47(7):947-54. doi:10.1007/s00256-018-2919-3.
22. 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. doi:10.1016/j.pmrj.2009.11.006.
23. Lee JC, Cha J-G, Kim Y, Kim Y-I, Shin B-J. Quantitative Analysis of Back Muscle Degeneration in the Patients With the Degenerative Lumbar Flat Back Using a Digital Image Analysis. Spine. 2008;33(3):318-25. doi:10.1097/BRS.0b013e318162458f.