Published in final edited form as:

J Manipulative Physiol Ther. 2020 January ; 43(1): 43–49. doi:10.1016/j.jmpt.2018.11.027.

Reliability of Human Lumbar Facet Joint Degeneration Severity assessed by magnetic resonance imaging

Joshua W. Little, DC, PhD,

Department of Surgery, Center for Anatomical Science and Education, School of Medicine, Saint Louis University, St. Louis, MO

Thomas Grieve, DC, MPH,

Department of Clinical Sciences, National University of Health Sciences, Lombard, IL

Joseph Cantu, DC

Department of Research, National University of Health Sciences, Lombard, IL

William B. Bogar, DC, Rudy Heiser, DC, Heather Miley, DC

Department of Diagnostic Imaging, National University of Health Sciences, Lombard, IL

Gregory D. Cramer, DC, PhD

Department of Research, National University of Health Sciences, Lombard, IL

Abstract

Objective: The purpose of this study was to determine the reliability of the assessment of lumbar facet joint degeneration severity by analyzing degeneration subscales using magnetic resonance imaging (MRI) in human subjects.

Methods: The reliability of articular cartilage degeneration, subchondral bone sclerosis, and osteophyte formation subscales of lumbar facet joint degeneration severity was assessed in MRI images from n=10 human subjects. Each scale was applied to n=20 lumbar facet joints (L4/5 level). Three examiners were trained. A first assessment of MRI images was provided by the examiners followed by a second assessment 30 days later. Intra-observer and inter-observer reliability were determined using percent agreement and the weighted kappa coefficient κ_w for paired comparisons and the overall kappa κ_o . The minimum threshold for reliability was set at moderate levels of agreement, $\kappa_w > 0.40$, based upon previous recommendations.

CORRESPONDING AUTHOR CONTACT INFORMATION: Joshua W. Little, DC, PhD, josh.w.little@gmail.com, 70 Winding Pine Trail Cary, North Carolina 27519, (314) 550-8732.

Contributorship: *Concept development* (provided idea for the research), JL, TG, GC, *Design* (planned the methods to generate the results), JL, TG, GC, *Supervision* (provided oversight, responsible for organization and implementation, writing of the manuscript), JL, TG, GC, *Data collection/processing* (responsible for experiments, patient management, organization, or reporting data), JL, TG, JC, WB, RH, HM, GC, *Analysis/interpretation* (responsible for statistical analysis, evaluation, and presentation of the results), JL, TG, GC, *Literature search* (performed the literature search), JL, TG, GC, *Writing* (responsible for writing a substantive part of the manuscript), JL, TG, GC, *Critical review* (revised manuscript for intellectual content, this does not relate to spelling and grammar checking), JL, TG, JC, WB, RH, HM, GC, *Other* (list other specific novel contributions)

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Results: The articular cartilage subscale had acceptable intra-observer ($\kappa_o=0.54$) and inter-observer ($\kappa_o=0.44$) reliability. Scales for subchondral bone sclerosis (intra-observer $\kappa_o=0.32$; inter-observer $\kappa_o=0.10$) and osteophyte formation (intra-observer $\kappa_o=0.26$; inter-observer $\kappa_o=0.26$) did not achieve acceptable reliability.

Conclusions: Of the three subcategories of lumbar facet joint degeneration, only articular cartilage degeneration demonstrated acceptable reliability. Subscales of lumbar facet joint degeneration should be considered independently for reliability before combining subscales for a global degeneration score. Due to the inherent difficulty assessing lumbar facet joint degeneration, the use of multiple examiners independently assessing degeneration with reliable scales and then coming to a consensus score upon any disagreements is recommended for future clinical studies.

MeSH

Osteoarthritis; Lumbar Spine; Zygapophyseal Joint; Facet Joint; Diagnostic Imaging; Reliability; Magnetic Resonance Imaging

Introduction

Lumbar facet joint, also known as zygapophyseal or Z joint, degeneration consistent with spinal osteoarthritis affects millions of individuals and can be associated with joint dysfunction and pain.¹ Assessment of lumbar spinal degeneration can be performed with standard radiography, computed tomography (CT), and magnetic resonance imaging (MRI).¹ Advanced imaging techniques such as CT and MRI offer distinct advantages over standard radiography including detection of less severe stages of degeneration, although all imaging techniques underestimate degeneration compared to the standard of histological facet joint analysis.¹⁻⁴ Previous studies assessing the reliability of grading the severity of osteoarthritic changes to the lumbar facet joint have been performed and reviewed.⁵ The reliability of assessing facet joint degeneration is typically lower than other joints, and because of these difficulties, recommendations have been made to conduct reliability studies on the methods to be used prior to their implementation in clinical studies.⁵ Although a few scales have shown acceptable reliability in imaging of human subjects for grading lumbar facet joint degeneration for CT, MRI, and radiographs,^{4,6,7} several other scales have been reported that were unreliable or did not assess reliability.^{5,8-11}

We previously reported a study of human lumbar facet joint degeneration severity with standard radiography and found acceptable reliability using a modified Kellgren method.⁶ However, the use of standard radiography is less sensitive and detects only the more advanced stages of facet joint degeneration compared to CT and MRI.¹⁻⁵ A recent movement in the field of osteoarthritis is to examine, in addition to the articular cartilage, the role of other joint tissues as a whole-joint disease in the induction, progression, and clinical presentation of joint degeneration.^{1,12} Degenerative changes to the lumbar facet joints can include pathological alterations in the articular cartilage, subchondral bone, synovium, synovial fold, and capsule.^{1,13} CT and MRI have distinct advantages in the detection of changes to multiple tissues of the lumbar facet joint. CT is more adept at revealing bone-related changes (e.g., subchondral sclerosis, osteophyte formation, joint space narrowing) while MRI is preferred to detect changes to articular cartilage, synovium, and bone marrow

spaces (e.g., articular cartilage thinning, synovitis, and bone marrow enlargement).^{1,4,5} Thus, the use of CT and MRI may be more advantageous to assess the severity of lumbar facet joint degeneration as a whole joint disease.

Previous grading system approaches in human subjects have used combinations of facet joint degeneration subcategories to produce an overall degeneration score.^{4–11,14} Those subcategories included articular cartilage degeneration with subchondral sclerosis and articular cartilage degeneration with osteophyte formation.^{4,5,7} Although those studies reported acceptable reliability for an overall degeneration score, they did not report the reliability of each subcategory. Other approaches used MRI in cadaveric specimens with dissected lumbar spinal segments to assess the reliability of each subcategory for articular cartilage, subchondral bone, and osteophyte formation.¹⁵ There was acceptable reliability for each subscale for the cadaveric specimens.¹⁵ The reliability of each subscale including the subcategories of articular cartilage, subchondral bone, and osteophyte formation has not yet been reported in human subject MRI for clinical studies. The purpose of this study was to examine the reliability of the subcategories of degeneration for articular cartilage, subchondral bone sclerosis, and osteophyte formation in lumbar facet joints using MRI in live human subjects. We hypothesized that using 4 point subcategory scales modified from previous advanced imaging studies in cadavers¹⁵ would be reliable in human subjects MRI studies based upon the recommended threshold for moderate agreement in imaging studies of the lumbar facet joint.⁵ Additionally, we hypothesized that because we were using MRI, the articular cartilage subcategory of degeneration would show the greatest levels of reliability.⁵ If reliable, these subcategories could be used to develop a global degeneration score, which would be advantageous for future clinical studies of lumbar facet joint osteoarthritis.

Methods

Subjects

All experiments were performed following approval and recommendation of the National University of Health Sciences Institutional Review Board. Using recruitment strategies previously successfully implemented by our research group,¹⁶ human subjects (n=10) were recruited from the greater Chicago metropolitan area, with emphasis on the western suburbs, as part of pilot studies for a larger scale clinical trial.¹⁷ Females (n=5) and males (n=5), mean age of 42.7 +10.2 years; (males 43.6, +10.2; females 41.4, +10.9), were recruited based upon inclusion and exclusion criteria as previously reported.¹⁷ Following informed consent, subjects underwent clinical examination and were subsequently imaged with MRI of the lumbar spine.

MRI Protocol

The lumbar spine was imaged and optimized to visualize the right and left L4/5 facet joints. The following MRI procedures were used for all scans obtained throughout the study. The subjects were positioned supine within the gantry of a Hitachi MRP 5000, 0.2-tesla MRI unit. A Hitachi quad-body coil #2 was positioned around the subject. Axial scans were from superior to inferior (TR 750, TE 25, FA 90°, 6 signal averages). The scan plane passed

through the inferior aspect of the intervertebral disc–superior vertebral body margin junction. The scans covered the entire region of the facet joint with five contiguous 3-mm-thick slices.

Degeneration Severity Scales

Degeneration was assessed on MRI images with subscales for categories of for articular cartilage degeneration, subchondral bone sclerosis, and osteophyte formation as previously reported.¹⁵ Each subcategory had a four point scale of progressive degenerative findings.

Articular cartilage degeneration subscale included Grade 1: articular cartilage is uniform, thick cartilage covering the entire articular surfaces (normal); Grade 2: articular cartilage covering the entire articular surfaces with irregular regions of erosion; Grade 3: incomplete cartilage coverage of the articular surface with erosion leaving bone exposed to the joint; Grade 4 articular cartilage is generally absent from the articular surfaces, except for trace amounts.

Subchondral bone sclerosis subscale included Grade 1: presence of cortical bone covering the articular processes (normal); Grade 2: cortical bone focally thickened (dense) on the articular processes; Grade 3: presence of mildly to moderately sclerotic/thickened cortical bone on the articular processes (mild to moderate); Grade 4: presence of thickened (dense) cortical bone covering >1/2 of the articular processes.

Osteophyte formation subscale included Grade 1: no osteophyte (normal); Grade 2: possibility or presence of a small osteophyte; Grade 3 presence of a mild to moderate amount of osteophyte formation and/or facet hypertrophy; Grade 4: presence of a large osteophyte.

Reliability Studies

Acquired images were selected by the authors (TG and WB) and neither were examiners in the reliability study. For the reliability study, images were selected to ensure representation of all grades within each subscale and that no overlying pathologies were present that would prevent examination and interpretation of facet joint degeneration. Three examiners included a DACBR certified chiropractic radiologist and 2 chiropractic radiology residents. Examiners graded left and right L4/L5 facet joint degeneration of 20 joints from 10 subject MRI scans. An MRI atlas showing examples of facet joint degeneration was constructed for each 4-point subscale including cartilage degeneration, subchondral bone sclerosis, and osteophyte formation. The three examiners were trained to use each subscale by receiving instruction through a series of two training sessions in the grading method for degenerative change. The atlas was provided to each of the 3 examiners at both training sessions and was available for reference during the examination of facet joints for the study. Each study scan and jacket was marked only with a random number generated for the study. Each MRI scan contained images of the left and right L4/5 facet joints that were determined by the facet joint study radiologist to show the facet joints to best advantage. To prevent potential examiner bias, the patient's clinical information was withheld. The examiners used the facet joint degeneration criteria to independently grade the facet joints of the 10 MRI scans; grading each joint using a 1–4 scoring system for each of the three subscales. The grading

was then repeated after 30 days. The second grading period was preceded by an additional training session. The examiners were blinded to the other observers' results and their previous scoring.

Statistical Analysis

Intra- and inter-observer paired agreements were assessed for each subscale using percent agreement and weighted kappa (κ_w) coefficients that were calculated using Medcalc (v18.5). Percentages are reported individually and as mean $\pm$ standard deviation (SD) and κ_w coefficients are reported with standard error (SE). An overall weighted kappa (κ_o), weighted to account for examiner variance with overall standard error (SE_o), was calculated to report the overall inter- and intra-observer agreements with 95% confidence interval (CI) for each subscale for all three examiners as previously described.¹⁸ Reliability was determined based upon the categorical ranges described by Bland and Altman: 0=poor, 0.01–0.20=slight, 0.21–0.40=fair, 0.41–0.60=moderate, 0.61–0.80=substantial, 0.81–1.0=almost perfect agreement.¹⁹ κ_w provides weighted values to account for the degree of disagreement between two observers and is the preferred method for evaluating ordinal data for reliability, as previously reported in imaging reliability studies of the facet joint osteoarthritis grading scales.^{4,5} An acceptable level of reliability was defined based upon previous recommendations for reliability of assessing lumbar facet joint degeneration requiring a score of $\kappa_w > 0.40$.⁵

Results

MRI images from human subjects (n=10) included the same number of females (n=5; 41.4 $\pm$ 10.9 years old) and males (n=5; 43.6, $\pm$ 10.2 years old) and were analyzed for lumbar facet joint degeneration subscale reliability.

Examination of the articular cartilage degeneration subscale (Figure 1, Table 1) revealed mean intra-observer agreements of 57% perfect agreement (75%, 60%, and 35%) and agreement within one grade was a mean of 98.3% (100%, 100%, and 95%). Overall intra-observer agreement was $\kappa_o = 0.54$ (95% CI=0.37–0.69); paired scores were examiner 1 $\kappa_w = 0.71$ SE=0.15, examiner 2 $\kappa_w = 0.52$ SE=0.13, and examiner 3 $\kappa_w = 0.39$ SE=0.16. Mean inter-observer perfect agreement was 47% (40%, 25%, and 75%) with mean agreement within one grade at 100%. Overall inter-observer agreement was $\kappa_o = 0.44$ (95% CI=0.29–0.58); paired scores were examiner 1 & 2 $\kappa_w = 0.45$ SE=0.11, examiner 2 & 3 $\kappa_w = 0.58$ SE=0.14, and examiner 1 & 3 $\kappa_w = 0.28$ SE=0.14. Examination of the subchondral bone subscale (Table 1) revealed mean intra-observer perfect agreement was 48% perfect agreement (65%, 50%, and 30%) with agreements within one grade at 88% (100%, 100%, and 65%). Overall intra-observer agreement was $\kappa_o = 0.32$ (95% CI=0.14–0.49); paired scores were examiner 1 $\kappa_w = 0.38$ SE=0.12, examiner 2 $\kappa_w = 0.38$ SE=0.19, and examiner 3 $\kappa_w = 0.08$ SE=0.20. Mean inter-observer perfect agreement was 33% (50%, 50%, and 15%) with agreements within one grade at 77% (100%, 65%, and 65%). Overall inter-observer agreement was $\kappa_o = 0.10$ (95% CI=–0.03–0.22); paired scores were examiner 1 & 2 $\kappa_w = 0.20$ SE=0.27, examiner 2 & 3 $\kappa_w = 0.05$ SE=0.08, and examiner 1 & 3 $\kappa_w = 0.19$ SE=0.11. Examination of the osteophyte subscale (Table 1) revealed mean intra-observer perfect

agreement was 57% (60%, 60%, and 50%) with agreement within one grade a mean of 93.3% (90%, 100%, and 90%) Overall intra-observer agreement was $\kappa_o = 0.26$ (95% CI=0.05–0.46); paired scores were examiner 1 $\kappa_w = 0.22$ SE=0.26, examiner 2 $\kappa_w = 0.32$ SE=0.13, and examiner 3 $\kappa_w = 0.10$ SE=0.23. Mean inter-observer perfect agreement was 38% (35%, 55%, and 25%) and mean agreement within one grade was 87% (85%, 95%, and 80%). Overall inter-observer agreement was $\kappa_o = 0.20$ (95% CI=0.07–0.33) paired scores were examiner 1 & 2 $\kappa_w = 0.27$ SE=0.11, examiner 2 & 3 $\kappa_w = 0.07$ SE=0.09, and examiner 1 & 3 $\kappa_w = 0.40$ SE=0.15. As the recommended threshold for facet joint degeneration scale reliability is a $\kappa_w > 0.40$,⁵ the percent agreement and κ scores show only the articular cartilage subscale to be reliable.

Discussion

The purpose of this study was to determine the reliability of grading the severity of lumbar facet joint degeneration using subcategories of articular cartilage degeneration, subchondral bone sclerosis, and osteophyte formation on MRI for human subjects in clinical studies. We found that of the three subcategories, the most reliable subscale for use on MRI in clinical studies was the articular cartilage degeneration. Subchondral bone and osteophyte formation subscales did not meet the recommended threshold for use in grading the severity of lumbar facet joint osteoarthritis. As such, we cannot recommend combining the currently studied subcategories to create a global score of facet joint degeneration severity for MRI. We would recommend that prior to combining subcategories into a global scale each subscale should be examined for reliability in MRI studies.

Our finding that articular cartilage degeneration was the most reliable to detect the severity of facet joint osteoarthritis was expected. MRI is advantageous for visualizing the non-osseous structures of the facet joint including articular cartilage.^{1,5} Additional MRI studies of human subjects in the future should consider assessing for pathological alterations in tissues such as the synovium and bone marrow spaces in addition to the articular cartilage, which may provide more advantageous subscales for better global assessment of degeneration severity with MRI. CT is the most advantageous for detecting the osseous features of the joint including the subchondral sclerosis and osteophyte formation that occur during spinal osteoarthritis.^{1,5} Perhaps the results of our present study would be reversed if applied to CT imaging. Additional studies could consider the use of both MRI and CT within the same subject to make direct comparisons related to the most advantageous comparisons for each imaging modality. The findings from our study support that the reliability of subscales for CT and MRI studies should be determined prior to using the combination of subscales in global or overall degeneration severity scores.

Previous studies using MRI approaches to study spinal osteoarthritis in human subjects have reported comparable levels of reliability using global scales that include combined categories for both cartilage and osteophytes as well as for all three subcategories of degeneration examined in this study.^{4,5} However, in these studies using global facet joint degeneration scales the reliability was performed on combined scales without first determining the reliability of each subcategory. Thus, the reliability of a global scale could be a result of one reliable subcategory. One MRI study in dissected lumbar spinal segments

from cadaveric specimens using the same scales reported in our study did demonstrate acceptable reliability for cartilage, subchondral sclerosis, and osteophyte formation subscales.¹⁵ When compared with MRI in living human subjects, the use of cadaveric specimens with dissected lumbar spinal segments and the ability to use MRI coils in closer proximity to the spine in these specimens likely optimizes the visualization of lumbar spinal structures allowing higher resolution.¹⁵ Our findings may also demonstrate that detecting lumbar facet joint degeneration from MRI in living human subjects is substantially more technically challenging than from cadaveric specimens.

LIMITATIONS

Limitations to this study include the inherent problem of detecting and classifying lumbar facet joint degeneration, regardless of imaging modality.⁵ These known difficulties have resulted in recommendations to accept the use of scales for research that can attain moderate levels of agreement.⁵ There are several approaches that could be used to enhance reliability such as using higher field strength (>1T) versus lower field strength (<1T) MRI and the use of expert examiners with equal number of years of experience. However, in studies assessing high (3T) versus low (0.25T) field strength MRI demonstrate excellent agreement, with high field strength providing little additional diagnostic benefit for grading lumbar spine degeneration.²⁰ Additionally, studies using training, atlases, expert examiners with over 12 years experience in musculoskeletal radiology and a simplified consensus defined scale still demonstrated similar levels of reliability as previously studies of facet joint degeneration.^{4–11,14} This likely reflects the inherent difficulty of grading lumbar facet joint degeneration in human subjects and the need for careful design and interpretation of facet joint imaging. To account for these limitations and in agreement with previous recommendations for imaging studies,²¹ we recommend that in future studies classifying the severity of facet joint degeneration on MRI, the articular cartilage subscale should be performed independently by two or more observers. Additionally, if there are disagreements in scoring, then the observers could also come together for a consensus score by meeting to form an agreement on discrepancies in scores. Data should be reported for independent observers to demonstrate intra- and inter-observer variance and also reported as a consensus score. Similar approaches were shown to increase the level of reliability when ordinal scales are used to grade degenerative changes in the spine using MRI.²²

Conclusions.

Here we demonstrate that the articular cartilage degeneration subcategory of lumbar facet joint degeneration may be the most advantageous for future clinical MRI studies in human subjects. The articular cartilage subscale showed the highest inter- and intra-observer agreement. The moderate levels of agreement are considered acceptable for assessment of facet joint degeneration. Our results indicate that using well-trained observers (examiners) making independent assessments can reach recommended moderate levels of agreement for cartilage degeneration with MRI. However, we recommend for future studies using the articular cartilage subscale that two or more observers independently examine images and report these scores and then come together during a “consensus meeting” to reach agreement

on those joints that were in disagreement to report a consensus score. Collectively, these results could be beneficial for use in future studies of spinal disorders and spinal therapies.

FUNDING SOURCES AND CONFLICTS OF INTEREST

Funding for this project was provided by the National Institutes of Health/National Center for Complementary and Integrative Health (grant # 2R01-AT000123-06S2). No conflicts of interest were reported for this study.

References

1. Gellhorn AC, Katz JN, Suri P. Osteoarthritis of the spine: the facet joints. *Nat Rev Rheumatol* 2013;9(4):216–24; doi:nrrheum.2012.199 [pii] 10.1038/nrrheum.2012.199. [PubMed: 23147891]
2. Lee JC, Cha JG, Yoo JH, Kim HK, Kim HJ, Shin BJ. Radiographic grading of facet degeneration, is it reliable? - a comparison of MR or CT grading with histologic grading in lumbar fusion candidates. *Spine J* 2012;12(6):507–14; doi:S1529-9430(12)00427-5 [pii] 10.1016/j.spinee.2012.06.003. [PubMed: 22770987]
3. Saal JS. General principles of diagnostic testing as related to painful lumbar spine disorders: a critical appraisal of current diagnostic techniques. *Spine (Phila Pa 1976)* 2002;27(22):2538–45; discussion 46; doi:10.1097/01.BRS.0000032127.87893.17. [PubMed: 12435989]
4. Weishaupt D, Zanetti M, Boos N, Hodler J. MR imaging and CT in osteoarthritis of the lumbar facet joints. *Skeletal Radiol* 1999;28(4):215–9. [PubMed: 10384992]
5. Kettler A, Wilke HJ. Review of existing grading systems for cervical or lumbar disc and facet joint degeneration. *European spine journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society* 2006;15(6):705–18; doi:10.1007/s00586-005-0954-y.
6. Little JW, Grieve TJ, Cramer GD, Rich JA, Laptook EE, Stiefel JP, et al. Grading Osteoarthritic Changes of the Zygapophyseal Joints from Radiographs: A Reliability Study. *Journal of manipulative and physiological therapeutics* 2015; doi:10.1016/j.jmpt.2014.12.009.
7. Pathria M, Sartoris DJ, Resnick D. Osteoarthritis of the facet joints: accuracy of oblique radiographic assessment. *Radiology* 1987;164(1):227–30. [PubMed: 3588910]
8. Butler D, Trafimow JH, Andersson GB, McNeill TW, Huckman MS. Discs degenerate before facets. *Spine* 1990;15(2):111–3. [PubMed: 2326704]
9. Coste J, Judet O, Barre O, Siaud JR, Cohen de Lara A, Paolaggi JB. Inter- and intraobserver variability in the interpretation of computed tomography of the lumbar spine. *Journal of clinical epidemiology* 1994;47(4):375–81. [PubMed: 7730862]
10. Grogan J, Nowicki BH, Schmidt TA, Haughton VM. Lumbar facet joint tropism does not accelerate degeneration of the facet joints. *AJNR American journal of neuroradiology* 1997;18(7):1325–9. [PubMed: 9282864]
11. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthritis. *Ann Rheum Dis* 1957;16(4):494–502. [PubMed: 13498604]
12. Hunter DJ, Felson DT. Osteoarthritis. *BMJ* 2006;332(7542):639–42; doi:332/7542/639 [pii] 10.1136/bmj.332.7542.639. [PubMed: 16543327]
13. Lewin T Osteoarthritis in Lumbar Synovial Joints. A Morphologic Study. *Acta orthopaedica Scandinavica Supplementum* 1964:SUPPL 73:1–112.
14. Carrino JA, Lurie JD, Tosteson AN, Tosteson TD, Carragee EJ, Kaiser J, et al. Lumbar spine: reliability of MR imaging findings. *Radiology* 2009;250(1):161–70; doi:10.1148/radiol.2493071999. [PubMed: 18955509]
15. Fujiwara A, Lim TH, An HS, Tanaka N, Jeon CH, Andersson GB, et al. The effect of disc degeneration and facet joint osteoarthritis on the segmental flexibility of the lumbar spine. *Spine (Phila Pa 1976)* 2000;25(23):3036–44. [PubMed: 11145815]
16. Cambron JA, Dexheimer JM, Chang M, Cramer GD. Recruitment methods and costs for a randomized, placebo-controlled trial of chiropractic care for lumbar spinal stenosis: a single-site pilot study. *J Manipulative Physiol Ther* 2010;33(1):56–61; doi:S0161-4754(09)00298-X [pii] 10.1016/j.jmpt.2009.11.002. [PubMed: 20114101]

17. Cramer GD, Cambron J, Cantu JA, Dexheimer JM, Pocius JD, Gregerson D, et al. Magnetic resonance imaging zygapophyseal joint space changes (gapping) in low back pain patients following spinal manipulation and side-posture positioning: a randomized controlled mechanisms trial with blinding. *J Manipulative Physiol Ther* 2013;36(4):203–17; doi:10.1016/j.jmpt.2013.04.003. [PubMed: 23648055]
18. Haas M Statistical methodology for reliability studies. *Journal of manipulative and physiological therapeutics* 1991;14(2):119–32. [PubMed: 2019821]
19. Bland JM AD. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1(8476):307–10. [PubMed: 2868172]
20. Lee RK, Griffith JF, Lau YY, Leung JH, Ng AW, Hung EH, et al. Diagnostic capability of low-versus high-field magnetic resonance imaging for lumbar degenerative disease. *Spine* 2015;40(6):382–91; doi:10.1097/BRS.0000000000000774. [PubMed: 25584942]
21. Levine D, Bankier AA, Halpern EF. Submissions to Radiology: Our Top 10 List of Statistical Errors. *Radiology* 2009;253:288–90.
22. Espeland A, Vetti N, Krakenes J. Are two readers more reliable than one? A study of upper neck ligament scoring on magnetic resonance images. *BMC medical imaging* 2013;13:4; doi:10.1186/1471-2342-13-4. [PubMed: 23327567]

Figure 1. Examples of grades of articular cartilage degeneration (C).

A: Grade 1: uniform, thick cartilage covering the entire “normal” articular surfaces (arrow).

B: Grade 2: cartilage covering the entire articular surfaces with irregular regions of erosion (arrows).

C: Grade 3: incomplete cartilage coverage of the articular surface, with erosion leaving bone exposed to the joint (arrows).

D: Grade 4: cartilage is generally absent from the articular surfaces, except for trace amounts; joint space decreased (arrow).

Table 1:

Reliability of lumbar facet joint degeneration subcategories

	Intra-Observer Reliability			Inter-Observer Reliability		
	Perfect Agreement ^I	Agreement Within 1 Grade ^I	K _o	Perfect Agreement ^I	Agreement Within 1 Grade ^I	K _o
Articular Cartilage	56.67 ± 20.21	98.33 ± 2.89	0.54 * (0.37–0.69)	46.67 ± 25.66	98.33 ± 2.89	0.44 * (0.29–0.58)
Subchondral Bone	48.33 ± 17.56	88.33 ± 20.21	0.32 (0.14–0.49)	33.33 ± 17.56	76.67 ± 20.21	0.10 (–0.03–0.22)
Osteophyte Formation	56.67 ± 5.77	93.33 ± 5.77	0.26 (0.05–0.47)	38.33 ± 15.28	86.67 ± 7.64	0.26 (0.07–0.33)

^I Agreement values are percentages ± standard deviation (SD) and κ_O coefficients are reported as overall average with 95% CI.

* K > 0.40: meets the minimum threshold for acceptable reliability for detecting lumbar facet joint degeneration with imaging studies.