Volume 5 • Issue 4 | 127

fortune Journals

JOURNAL OF SPINE RESEARCH AND SURGERY

ISSN: 2687-8046

Research Article

Radiological Classification for Degenerative Lumbar Spine Disease: A Literature Review of the Main Systems

Marcelo Molina[1,2,3,*] and Sebastián Vial[1,4]

Abstract

Study Design: Systematic review

Objectives: Performed a systematic review of available lumbar spinal degenerative disease classifications.

Methods: We performed a systematic literature review search for papers that proposed or described radiological classification systems for degenerative lumbar spine disease, such as lumbar disc herniation, facet joint arthritis, spondylolisthesis, and lumbar stenosis. The literature was performed in MEDLINE and EMBASE, limited to English articles published from 1980 to the present. The reliability tests of the reviewed articles were assessed with the “Intraclass Correlation Coefficients” (ICC) and “Cohen's Kappa coefficient” (k).

Results: We found 1873 articles. A total of 64 articles were reviewed, identifying 31 radiological classification systems. We found 7 classifications for degenerative disc disease, 7 for disc herniation, 7 for facet joint osteoarthritis, 8 for degenerative spinal stenosis, and 2 for degenerative spondylolisthesis. Of the 31 systems found, 24 had interrater agreement studies. The clinical orientation of the classification was analyzed when appropriate.

Discussion: Reliability studies play a crucial role in evaluating a classification system as they enable reproducibility among evaluators, thereby fortifying the system. Classifications should not only be endorsed based on their validation and reliability studies, but it is also crucial to assess their feasibility for practical implementation in clinical settings.

Conclusions: A classification system should have a reliability with Kappa or ICC over 0.60 to be recommended. It should provide a clinical orientation to make therapeutic decisions and form part of a guideline. Continued research on classification development is essential to improve systems, enhancing their clinical utility and bolstering their reliability.

Keywords: spine; degenerative; lumbar; classification; disc herniation; spondylolisthesis; facet joint osteoarthritis; spinal; stenosis; literature; review

Introduction

Degenerative changes in the lumbar spine, considered a natural part of aging, typically commence early, often between the 2nd and 3rd decade of life [1]. These changes encompass all the components of the vertebral unit, including the intervertebral disc, the facet joints, their respective ligaments, and the adjacent vertebrae.

Affiliation:

[1]Orthopaedic Surgery Department, Spine Unit, Instituto Traumatológico, Dr. Teodoro Gebauer, Santiago, Chile

[2]Orthopaedic Surgery Department, Spine Unit, Clínica Alemana, Santiago, Chile

[3]Universidad Finis Terrae, School of Medicine, Chile

[4]Universidad de Chile, School of Medicine

***Corresponding author:**
Marcelo Molina, Orthopaedic Surgery Department, Spine Unit, Instituto Traumatológico, Santiago, Chile

Citation: Marcelo Molina , Sebastián Vial. Radiological Classification for Degenerative Lumbar Spine Disease: A Literature Review of the Main Systems. Journal of Spine Research and Surgery. 5 (2023): 127-138

Received: December 06, 2023
Accepted: December 11, 2023
Published: December 29, 2023

Molina M and Vial S., J Spine Res Surg 2023
DOI:10.26502/fjsrs0068

The distinction between normal and pathological degeneration is not always clearly defined. There is a gradual degenerative process of the spine that in some individuals progresses to pathological changes that can lead to symptoms. Numerous classification systems aim to quantify the degree of spine degeneration, delineating the threshold between normal and pathological changes. The severity of radiological degeneration is expected to correlate with clinical symptoms and functional status [2]. An ideal classification should allow to definition of the degree of severity of the pathology, provide a common language among health care professionals, establish a prognosis, and guide treatment [3]. Unfortunately, the relationship between pathological degeneration and symptoms does not always correlate closely. This paradox is evident in some patients presenting with severe radiological degeneration and mild symptoms, contrasting with others presenting without significant degeneration but with severe clinical impairment.

Despite the above limitations, imaging remains a primary tool in clinical practice for assessing spinal pathology, offering various classifications based on different imaging modalities described in the literature. We conducted a systematic review of available classifications for lumbar spinal degenerative disease, to identify those demonstrating the highest clinical utility, through a strong clinical correlation and reproducibility.

Material and Methods

We performed a systematic literature review search for papers that proposed or described a radiological classification system for degenerative lumbar spine disease, including lumbar disc herniation, facet joint arthritis, spondylolisthesis, and lumbar stenosis.

Information Search

The search was performed by 2 investigators, an experienced spinal surgeon, and an orthopedic resident. The literature search was carried out in three stages. The first one consisted in searching in MEDLINE and EMBASE databases using diverse combinations of the following MeSH terms: "classification", "diagnostic imaging", "computed tomography scan", "magnetic resonance imaging", "lumbar vertebrae", "Intervertebral disc", "intervertebral disc degeneration", "Intervertebral Disc Displacement", "spine osteoarthritis", "zygapophyseal joint", "spinal stenosis", "spondylolisthesis". The search was restricted to English, with articles published from 1980 to date. Next, we expanded the search by exploring the reference list of selected articles and utilizing the “related articles” function within the search engine. Lastly, only articles meeting our criteria were selected.

Inclusion criteria

We included clinical papers describing classifications of lumbar degenerative diseases based on radiographic, CT, and/or MRI evaluation. Preference was given to those with reliability assessments.

Evaluation of the classification systems:

The reliability tests for the reviewed articles were primarily assessed using the “Intraclass Correlation Coefficients” (ICC) and “Cohen's Kappa coefficient” (k), following "Landis & Koch criteria [4,50] (refer to Table 1).

A good classification reliability is generally accepted with a Kappa index of > 0.60, with at least substantial agreement [2]. However, for facet joint degenerative pathology -which is considered more difficult to classify- Kappa coefficients or ICC >0.40 (at least moderate agreement) is considered acceptable for evaluation [2].

Results

Initially, 1873 articles were identified in the databases. Articles were selected according to their title and abstract, subsequently, bibliographic references and related articles in search engines were reviewed. A total of 64 articles were reviewed, identifying 31 imaging-based classification systems for lumbar spine degenerative pathology that adjusted to our selection criteria.

Lumbar Disc Degeneration Classifications Systems:

In our literature search, we found seven imaging systems for lumbar disk degeneration, all with reliability evaluation. Five were based on MRI findings and two were based on plain radiographs (see Table 2). However, because MRI is the current gold standard, we only included MRI-based classifications.

Only two underwent both intra- and inter-rater reliability tests, showing a Kappa coefficient > 0.60. The classification proposed by Pfirrmann et al. [7] and its subsequent modification by Griffith et al.8 exhibit greater values in their reliability tests. The Pfirrmann et al. MRI-based classification categorizes lumbar disc degeneration into 5 degrees based on signal intensity, disc structure, and differentiation between the nucleus and the disc ring [7]. Griffith et al.'s modification increases this to eight, enhancing discrimination, especially between disc degeneration in elderly subjects. In the original

Table 1: Agreement Measures for categorical data according to the criteria published by "Landis & Koch [4,50].

Kappa Statistic	Strength of Agreement
< 0.00	Poor
0.00 – 0.20	Sligth
0.21 – 0.40	Fair
0.41 – 0.60	Moderate
0.61 – 0.80	Substantial
0.81 – 1.00	Almost Perfect

Molina M and Vial S., J Spine Res Surg 2023
DOI:10.26502/fjsrs0068

Pfirrmann classification, there were more than 87% of lumbar intervertebral discs graded as either III or IV in this age group, without substantial difference between them [8].

The "Tufts Classification for lumbar disc degeneration", created by Riesenburger et al. classifies degenerative disc disease in 6 degrees (Grades 0 to 5) depending on the score obtained according to the variables of disc brightness and structure, Modic changes, high-intensity zones (HIZ) and disk height. In their reliability studies, moderate to excellent intra-rater agreement (k = 0.53 - 0.94) and substantial inter-rater agreement (k = 0.66 - 0.77) were demonstrated for all the variables except for HIZ, which showed moderate agreement [9]. Later, Burke et al. modified this classification system eliminating the HIZ variable. At the same time, he performed reliability tests involving evaluators from different specialties (2 neuroradiologists and 2 neurosurgeons), to assess inter-specialty reproducibility. The interrater agreement was moderate (k = 0.465 - 0.576) and the intrarater agreement was moderate to substantial (k = 0.523 - 0.649) [10].

Lumbar Disk Herniation Classifications Systems:

Seven classification systems related to LDH were identified (refer to Table 3). Four articles [12], [13], [15], and [46] underwent reliability evaluation. The classification suggested by Ahn et al. [12] uses MRI to grade the sagittal migration of LDH from 1 to 6, depending on the direction and distance from the disc space, displaying substantial intra and inter-observer agreement (refer to Table 3). The Michigan State University (MSU) classification system, developed by Mysliwiec et al. [15] using weighted-T2 axial MRI images, categorizes the LDH in 3 levels (from anterior to posterior: 1, 2, and 3) and the medial-lateral location in 4 levels (A = central; AB = paracentral; B = lateral recess; C = far lateral). The Kappa coefficient has almost perfect values for both the interrater (weighted k: Grade = 0.934; Location = 0.904) and intrarater (Weighted Kappa: Grade = 0.883; Location = 0.808) agreement. Halldin et al. [13], through CT and MRI, present a graduation system for LDH distribution and size across multiple planes. They did not use the Kappa coefficient or ICC for the evaluation of intra and interobserver

Table 2: Summary of the Lumbar Disc Degeneration Classifications Systems included in our review

Method	Grading system	Intraobserver reliability	Interobserver reliability	Comments	Summary
MRI	Schneiderman et al [6]	Not determined in original article	Not determined in original article	2 radiologist	4 grades of signal intensity (normal, intermediate los, marked loss or absent signal)
	Pfirrmann et al., 2001 [7]	Kappa = 0.84 - 0.90	Kappa = 0.69 - 0.81	1 orthopedic surgeon, 2 radiologists	5 grades (grade I to V)
	Modified Pfirrmann (Griffith et al., 2007) [8]	Kappa = 0.79 -0.91	Kappa = 0.65 - 0.67	3 radiologists	8 grades (grade 1 to 8)
	Tufts Lumbar Degenerative Disc Classification (Riesenburger et al.,2015) [9]	Average Kappa: Observer 1: Disc brightness: 0.78; Modic changes: 0.83; HIZ: 0.75; disc height: 0.94; Total grade: 0.71 Observer 2: Disc brightness: 0.66; Modic changes: 0.71; HIZ: 0.53; disc height: 0.54; Total grade: 0.58	Average Kappa: Disc brightness: 0.69; Modic changes: 0.66; Disc height: 0.77; HIZ: 0.41; Total grade: 0.57	2 neurosurgeons	Point system classification, include 4 radiographics indicators (Disc structure and brightness, modic changes, HIZ, disc height) . 6 grades (Grade 0 to 5)
	Modified Tufts Lumbar Degenerative Disc Classification (Burke et al., 2015) [10]	Kappa = 0.523 - 0.649	Kappa = 0.465 - 0.576.	2 neurosurgeons, 2 neuroradiologists. Exclude HIZ variable	Point system classification include 4 radiographics indicators (Disc estructure and brightness, modic changes). 6 grades (Grade 0 to 5)

Molina M and Vial S., J Spine Res Surg 2023
DOI:10.26502/fjsrs0068

Plain radiography	Lane et al. 1993 [5]	Average ICC/Kappa = 0.90 (Narrowing: 0.92; osteophytes: 0.96; Sclerosis: 0.59)	Average ICC/Kappa = 0.93 (Narrowing: 0.95; osteophytes: 0.91; Sclerosis 0.55)	3 examiners	Grading system based from the presence and severity of individual features
	Madan et al., 2003 [11]	Not determined in original article	Weighted Kappa = 0.351–0.673	5 examiners (2 orthopaedic surgeons, 2 radiologists and a spine nurse practitioner)	4 grades (None, Mild, Moderate and Severe)

HIZ = High intensity zone

Table 3: Summary of Lumbar Disc Herniation Classifications Systems included in our review

Method	Grading system	Intraobserver reliability	Interobserver reliability	Comments	Summary
MRI	Lee et al., 2007 [14]	Not determined in original article	Not determined in original article	Try to provide appropriate surgical guideline of PELD for migrated disc herniation.	4 zones depending on the direction and distance from disc space.
	Clasificación de “Michigan State University” - MSU (Mysliwiec et al., 2010) [15,44]	Not determined in original article * Weighted Kappa: - Degree: 0.934 - Location: 0.904	Not determined in original article * Weighted Kappa: - Degree: 0.883 - Location: 0.808	Try to correlate symptoms and images findings *1 spine specialist and 1 radiologist	The size of LHD is expressed as “1,2,3” and the location is expressed as “A, AB, B, C”
	Ahn et al., 2017 [12]	Average Kappa: - Reader 1 = 0.827 - Reader 2 = 0.620	Average Kappa: - 1st evaluation = 0.737 - 2nd evalution = 0.657	2 radiologists	6 grades of disc migration in the sagittal plane depending on the direction and distance from disc space
	Zhu et al., 2023 [46]	Intra-observer - Reader 1 (first) vs. reader 1 (second) 0.734 - Reader 2 (first) vs. reader 2 (second) 0.617	Inter-observer - Reader 1 (frst) vs. reader 2 (frst) 0.748 - Reader 1 (second) vs. reader 2 (second) 0.639	2 radiologists	4 types (1 to 4) according to the morphology of the LDH
MRI & CT	Wiltse et al.,1997 [16]	Not determined in original article	Not determined in original article	12 pysicians	5 grades of LDH size (1 to 5)
	Halldin et al., 2007 [13]	Not determined in original article	Not determined in original article	Try to correlate clinical and images findings Reliability tests are not calculated with kappa coefficient or ICC	Point system classification. The transverse plane was divided into 4 sectors each side, the sagittal plane was divided in 4 sectors and longitudinal distribution was divided in 3 levels
	Hao et al., 2017 [17]	Not determined in original article	Not determined in original article	Clinical-radiological classification 3 examiners	Point system classification. Types from I to V. Type III was subclassified into A, B and C.

PELD = Percutaneous endoscopic lumbar discectomy
* Intra and inter-observer weighted kappa coefficient values according to Zhu et al. [44]

Molina M and Vial S., J Spine Res Surg 2023
DOI:10.26502/fjsrs0068

agreement, this makes reliability studies not comparable with other classifications. Zhu et al., in 2023, introduced an MRI-based LDH classification system outlining four types and recommending a surgical strategy for each one, demonstrating good inter- and intra-observer agreement [46].

Lumbar Facet Joint Osteoarthritis Classifications:

Eight imaging systems of lumbar facet degeneration were found (refer to Table 4). Given the difficulty of evaluating degenerative pathology on facet joints, regardless of imaging modality, a Kappa coefficient or ICC> 0.40 is considered acceptable [2]. Only four classifications -Pathria et al. [18], Weishaupt et al. [19], Fujirawa et al. [20], and Little et al. [21] had reliability studies meeting these criteria. Pathria et al.'s classification using plain radiographs was excluded due to poor inter-observer agreement (k = 0.26) [18].

Pathria et al. in 1987 proposed a facet osteoarthritis severity classification system based on CT, categorizing it into 4 grades (from grade 0 to 3). They only evaluated the interrater agreement, which was k = 0.46 [18]. Fujirawa et al. building upon Pathria et al.'s work, introduced an MRI-based classification system with 4 severity degrees exhibiting substantial inter-rater agreement (k = 0.636). Stieber et al. performed new reliability studies for both classifications, showing different values compared to the original articles (refer to Table 4) [25]. This could be due to the difficulty in assessing the facet joint, independent of the imaging methodology.

Little et al.'s modification of Kellgren's classification 2015, uses radiographs to grade the severity of osteoarthritis by 5 degrees (from 0, no osteoarthritis, to 4, advanced osteoarthritis) with moderate to substantial inter-observer agreement (weighted average Kappa = 0.63) and moderate intra-observer agreement (weighted Kappa = 0.42 and 0.54) [21].

The Weishaupt et al. classification system has the best reliability studies among the identified systems. Ranging from grade 0 (normal facet joint space) to grade 3 (narrowing of the joint space and/or large osteophytes and/or severe hypertrophy of the joint process and/or severe subarticular erosion and/or subchondral cysts). Furthermore, it shows an adequate intra -and inter-observer agreement in both CT and MRI assessments (refer to Table 4) [19].

Degenerative Lumbar Spinal Stenosis Classifications:

Eight classification systems were found (refer to Table 5), all based on MRI.

The system proposed by Lurie et al. grades the severity of lumbar spinal stenosis in three areas: central, lateral recess, and foramina, and also evaluates root impingement. They defined it as "mild stenosis" if the decrease in area is ≤1/3 of the normal area, "moderate" if the compromise is between 1/3 and 2/3 of the normal area, and "severe" if the compromise is >2/3 of the normal area. For central stenosis, this classification shows an almost perfect intra- and inter-

Table 4: Summary of the Lumbar Facet Joint Osteoarthritis Classifications Systems included in our review

Method	Grading system	Intraobserver reliability	Interobserver reliability	Comments	Summary
Plain radiography	Kellgren modificada (Little et al., 2015) [21]	Weighted Kappa: Observador 1 = 0.42 Observador 2 = 0.54	Weighted Kappa = 0.63 (0.57, 0.60 y 0.68)	3 radiologists	5 grades (grade 0 to 4)
CT	Pathria et al., 1987 [18]	Not determined in original article *Kappa = 0.52 and 0.51.	Kappa = 0.46 *Kappa = 0.33 and 0.45	Two radiologists	4 grades (grade 0 to 3)
	Butler et al., 1990 [22]	Not determined in original article	Not determined in original article	Without reliability studies	2 grades (“normal”, “degenerative”)
	Coste et al., 1994 [23]	Kappa: Right facet joint = 0.16 (0.04-0.26) Left facet joint = 0.16 (0.06-0.27)	Kappa: Right facet joint = 0.03 (-0.16-0.18) Left facet joint = -0.01 (-0.13-0.11)	2 radiologists and 2 rheumatologists	2 grades (grade 1 and 2)
MRI	Grogan et al., 1997 [24]	Not determined in original article	Not determined in original article	Without reliability studies	Articular cartilage and sclerosis grade, each one with 4 grades (grade 1 to 4)
	Fujirawa et al., 1999 [20]	Not determined in original article *Kappa = 0.36 and 0.26	Kappa = 0.636 *Kappa: 0.22 and 0.10	2 orthopaedic surgeons	4 grades (grade 0 to 3)
MRI & CT	Weishaupt et al., 1999 [19]	Weighted Kappa: Examiner 1 (MRI/CT): 0.70/0.70 Examiner 2 (MRI/Ct): 0.76/0.77	Weighted Kappa: MRI: 0.41 CT: 0.60	2 radiologists	4 grades (grade 0 to 3)

* Reliability studies by Stieber et al [25].

Molina M and Vial S., J Spine Res Surg 2023
DOI:10.26502/fjsrs0068

rater agreement (k = 0.82) and substantial (Kappa: 0.73), respectively. However, for lateral recess stenosis, foraminal stenosis, and root impingement, they found only a moderate inter-rater agreement (refer to Table 4) [26].

Schizas et al. graded the severity of central and lateral recess stenosis by dural sac morphology on MRI, without using any specific measurement tools. The reliability evaluation was moderate for interobserver agreement (k = 0.44) and substantial for intraobserver agreement (k = 0.65) [27]. They graduated stenosis from A to D: Grade A corresponded to a mild stenosis or no stenosis and grade D was an "extreme stenosis". Grade A was subclassified into four specific subtypes of the distribution of the lumbar roots in the dural sac. Furthermore, they identified an association between grades C and D with a greater probability of failure of conservative treatment [27].

Table 5: Summary of the Degenerative Lumbar Spinal Stenosis Classifications Systems.

Method	Grading system	Intraobserver reliability	Interobserver reliability	Comments	Summary
MRI	Lauri et al., 2008 [26]	Kappa: Central stenosis = 0.82 (0.78–0.87) Subarticular stenosis = 0.75 (0.69–0.81) Foraminal stenosis = 0.77 (0.72–0.82) Root impingement = 0.76 (0.68–0.83)	Kappa: Central stenosis = 0.73 (0.69–0.77) Subarticular stenosis = 0.49 (0.42–0.55) Foraminal stenosis = 0.58 (0.53–0.63) Root impingement = 0.51 (0.42–0.59)	3 radiologists and 1 orthopedic surgeon	Severity rating on 4 grades Grading in 3 thirds ("none", "mild", "moderate", "severe")
	Schizas et al., 2010 [27]	Average Kappa = 0,65	Average Kappa = 0,44	Try to correlate symptoms and images findings 1 radiologist, 1 spinal surgeon and 2 orthopedic physicians	Grading stenosis from A to D. Grade A was subclassified into 1, 2, 3 and 4 according which the rootlets were disposed.
	Lee Guen et al., 2011 [28]	Kappa = 0.863 - 0.900	ICC = 0.730 – 0.953	4 radiologist Park et al.30 correlate symptoms and images findings	4 grades (Grade 0 to 3)
	Wildermuth et al., 1998 [30]	Not determined in original article	Average Kappa: 0.62	2 examiners	4 grades (grade 1 to 4)
	Lee et al., 2010 [31]	Kappa: L3–L4: right = 0.883, left = 1.00; L4–L5: right = 0.957, left = 0.885; L5–S1: right = 0.800, left = 0.905	Kappa: L3–L4: right = 1.0, left = 0.905; L4–L5: right = 0.929, left = 0.942; L5–S1: right = 0.919, left = 0.909	2 radiologists	4 grades (grade 0 to 3)
	Özer et al., 2022 [45]	Kappa values for stable types: I, 1.0; II, 1.0; III, 0.948; IV, 1.0; Kappa values for unstable types: I, 0.977; II, 0.982; III, 0.972; IV, 1.0.	Kappa values for stable types: I, 0.895; II, 0.939; III, 0.917; IV, 0.945 Kappa values for unstable type: I, 0.926; II, 0.919; III, 0.924; IV, 0.907.	1 neurosurgeon and 1 neuroradiologist	2 types of foraminal stenosis, stable and unstable. Each one with 4 subgrops (Type I to IV)
	Pfirrman et al., 2004 [47]	Kappa = 0.72–0.77	Kappa = 0.62–0.67	1 spinal radiology fellowship–trained orthopedic surgeon and 2 radiologists	four-grade scale based on compromise of the traversing nerve root
	Miskin et al., 2021 [48]	Not determined in original article	Kappa = 0.323 (0.255–0.392)	2 spine neurosurgeons, 2 spine orthopedic surgeons, 2 physiatrists, 1 musculoskeletal radiologist	Three-grade scale based on compromise of the traversing nerve root

Citation: Marcelo Molina, Sebastián Vial. Radiological Classification for Degenerative Lumbar Spine Disease: A Literature Review of the Main Systems. Journal of Spine Research and Surgery 5 (2023): 127-138

Molina M and Vial S., J Spine Res Surg 2023
DOI:10.26502/fjsrs0068

Another classification of degenerative central lumbar spinal stenosis, based on dural morphology by MRI and exhibiting good inter-observer agreement is described by Lee Guen et al. This system comprises four grades; ranging from grade 0 (the absence of stenosis) to grade 3, implying a severe stenosis with all the lumbar roots seen as a lump in the MRI [28]. Park et al. conducted a reliability study on Lee Guen et al. 's classification, revealing substantial inter-observer agreement values (k= 0.78). Additionally, they established an association between grade 0 with the absence of neurological manifestations and grade 3 with the presence of neurological manifestations [29].

For lateral recess stenosis, Pfirrman et al [47] in 2004 developed a grading system that described a four-grade scale based on compromise of the traversing nerve root (no compromise, contact of nerve root, deviation of nerve root, and compression of the nerve root). In 2021, Miskin et al [48] simplified this classification into a three-grade scale (normal, contact of the nerve root without compression, and compression of the nerve root). For lateral recess stenosis, Pfirmann et al. [47] reported an inter-reader agreement of 0.62–0.67 among three readers including one spine surgeon and two radiologists. Nevertheless, the modified classification had a fair agreement (k =0.323).

For foraminal stenosis, Wildermuth et al [30] in 1998 and Lee et al [31] in 2010 proposed an MRI classification with substantial and almost perfect Kappa coefficient values, respectively, for intra and interobserver agreement (see Table 5). Both consisted of 4 degrees, from the absence of foraminal stenosis to complete obliteration. It should be noted that the classification of Lee et al. is more precise in the definition of degrees. In 2022, Özer et al [45] proposed a new MRI classification and a treatment algorithm. They divided the foraminal stenosis into two groups: “stable” and “unstable”. In stable stenosis, the disc and annulus are calcified and facet joints are hypertrophic and degenerated. In unstable stenosis, there is a degenerative and mobile intervertebral disc. Each group has 4 subgroups about cause and type of compression. Then, they proposed a treatment for each subgroup. The classification has nearly perfect interobserver and intraobserver Kappa coefficient values (see Table 5).

Degenerative Spondylolisthesis Classifications:

Classically, spondylolisthesis has been classified according to the system proposed by Meyerding [32]. However, this classification is not specific for degenerative spondylolisthesis. We found only 2 imaging classifications for degenerative spondylolisthesis were found (refer to Table 6).

The “Clinical and Radiographic Degenerative Spondylolisthesis” (CARDS) classification system by Kepler et al. used static and dynamic plain radiographs, classifying the pathology with almost perfect interrater agreement (k = 0.82). It consists of 4 “radiographic types” (from A to D) based on disc collapse, anterior translation, and the presence of segmental kyphosis. Added to the above is the clinical variable modifying lower limb pain (0 = absent, 1 = unilateral, and 2 = bilateral) [33].

Gille et al. and the French Society for Spine Surgery, proposed a classification system using anteroposterior and lateral total spine radiograph [34], with almost perfect intra and interobserver agreement, with Kappa coefficients 0.89 and 0.82, respectively [35]. It should be noted that this system derives from the “Spinal Deformities in Adults Classification” by Schwab et al [36], which has an almost perfect (k = 0.87) and a substantial (k = 0.75) intra and interrater agreement, respectively. They classify spondylolisthesis into 5 types according to the calculations of the following variables: segmental lordosis, lumbar lordosis, pelvic incidence, pelvic tilt, and the vertical sagittal axis.

Table 6: Summary of the Degenerative Spondylolisthesis Classification Systems included in our review

Method	Grading system	Intraobserver reliability	Interobserver reliability	Comments	Reference
Plain radiography	Clasificación CARDS (Kepler et al, 2015) [33]	Kappa = 0.83 (0.77 - 0.89)	Kappa = 0.82 (0.74 - 0.90)	Clinical-radiological classification 5 fellowship trained spinal surgeons and a spine fellow	4 radiographic types (from A to D) plus a “leg pain modifier” (0 = absent; 1 = unilateral; 2 = bilateral)
	Gille el al., 2014 [34,35]	Not determined in original article **Kappa: 0.89	Not determined in original article **Kappa: 0.82	1 senior orthopedic surgeon and 2 orthopedic senior residents	5 types (type 1 to 5)

** Reliability studies by Ghailane et al [35]

Citation: Marcelo Molina, Sebastián Vial. Radiological Classification for Degenerative Lumbar Spine Disease: A Literature Review of the Main Systems. Journal of Spine Research and Surgery 5 (2023): 127-138

Molina M and Vial S., J Spine Res Surg 2023
DOI:10.26502/fjsrs0068

Discussion

In our systematic literature review, we identified 31 imaging-based classification systems for lumbar spine degenerative disease. This abundance of classifications underscores the intricate nature of the spine's functional unit, comprising various elements, including intervertebral discs, vertebral endplates, facet joints, and ligaments.

The different classification systems use various terminologies to describe the varying degrees of severity of lumbar degeneration. The evaluated classifications can use Arabic numerals (e.g., 1, 2, 3, etc.), Roman numerals (e.g., I, II, III, IV, etc.), letters (e.g., A, B, C, etc.) and qualitative terms (e.g., "mild," "moderate," "severe"). This variety may lead to confusion when comparing different grades. It should be mentioned that all the classifications, in their initial grades, describe the absence of degenerative pathology or mild changes. As graduation levels increase, more advanced stages of the disease are described, demonstrating a logical and evolutionary progression.

Reliability studies, particularly inter-observer agreement, play an important role in evaluating a classification system as they enable reproducibility among various evaluators, thereby fortifying the system's credibility. Typically, reliability is assessed using statistical tools like the Kappa coefficient or ICC. Among the 27 systems identified, 20 underwent inter-rater agreement studies. It is crucial not only to consider the validation and reliability of classifications but also their implementation in clinical practice. Simplicity and precision in the description of the systems are essential.

We identified one article by Kettler et al. [2], which assesses classifications for degenerative spine diseases, focusing on cervical and lumbar disc and facet joint degeneration, with an emphasis on their reliability studies. Their conclusions suggest preferred classifications based on statistical measures such as kappa or ICC values. In our article, we reviewed only image-based classifications of lumbar degenerative spine disease and include disc and facet joint degeneration, LDH, degenerative stenosis, and degenerative spondylolisthesis. However, we reviewed the lack of sufficient studies correlating them with clinical outcomes for prognosis. The choice of classification ultimately relies on the surgeon's expertise and experience.

Lumbar Disc Degeneration Classifications Systems

The gold standard image modality to assess disc degeneration is MRI. The system proposed by Pfirrmann et al. in 2001, evaluates lumbar disc degeneration using MRI and a simple algorithm to discriminate between 5 degrees. Moreover, it has an almost perfect and substantial intra and interrater agreement, respectively. Subsequent evaluations of its reliability showed almost perfect values of the Kappa coefficient and ICC, with no differences between specialties (radiologists versus spine surgeons) [38].

None of the other classifications found had the simplicity and inter-rater agreement reproducibility of Pfirrmann et al.'s classification. The modification made by Griffith et al. [8], allowing for more precise grading of degenerative lumbar disease across 8 levels, its complexity might limit its practicality in clinical settings."

Lumbar Disk Herniation Classifications Systems

In 2001, Fardon and his colleagues published an article detailing a consensus reached by members of the "North American Spine Society", the "American Society of Spine Radiology" and the "American Society of Neuroradiology" regarding the nomenclature and classification of lumbar disc disease38. This classification allows the characterization of LDH based on its morphology and location. An updated version by the same author was published in 2014 [39], proposing diagnostic categories for normal and pathological variations of LDH.

The MSU classification system [15] classifies the location and grade of the LDH with an almost perfect inter and intrarrater agreement [44]. Furthermore, there exists an association between the "MSU-B" LDH with greater severity of facet osteoarthritis. It should be noted that this system only evaluates LDH in the axial plane, not in the sagittal plane. However, the MSU classification is easy to apply in clinical practice with a very good level of agreement.

Another simple and easy system to use is Zhu et al. 's classification, which defines 4 types of LDH morphology and suggests a surgical strategy for each. However, de inter- and intra-observer agreement was only "good" [46].

Lumbar Facet Joint Osteoarthritis Classifications Systems

Only four imaging-based classifications identified for facet osteoarthritis meet the recommended international literature standards for Kappa coefficient or ICC values (>0.40) [2]. These encompass classifications utilizing plain radiographs (Little et al. [21]), CT scans (Pathria et al. [18]), MRI (Fujirawa et al. [20]), and a combination of CT and MRI (Weishaupt et al. [19]). It is known that the best image to evaluate the vertebral unit is MRI, therefore we recommend classification systems that use this imaging modality to evaluate the facet joins. Both Fujirawa et al. and Weishaupt et al. employ MRI to evaluate facet osteoarthritis, demonstrating adequate interobserver agreement. However, Weishaupt et al. 's classification offers a more detailed description between degrees, enhancing the precision of grading.

The reliability studies conducted by Weishaupt et al. [19] and Berg et al. [40] recommend the use of CT and MRI for assessing facet joint osteoarthritis. Although evaluation of the bony component (facet osteophytes and hypertrophy) is better with CT, results from CT and MRI are not significantly different.

Molina M and Vial S., J Spine Res Surg 2023
DOI:10.26502/fjsrs0068

Degenerative Lumbar Spinal Stenosis Classifications Systems

The best imaging modality to evaluate spinal stenosis is MRI due to its superior ability to identify non-bony components. The eight classifications identified use MRI to determine the severity of lumbar stenosis.

Fardon et al.'s classification offers a precise method to determine the location of spinal stenosis. It divides the site of compression into different zones based on anatomical limits. In the axial plane, these zones are the "central area", "lateral recess or subarticular area", "foraminal area", and "extraforaminal area". In the sagittal plane, these zones are the "infra pedicular level", "pedicular or disc level", and "supra pedicular level".

The severity of the stenosis, regardless of its location, could be properly assessed using the Lurie et al. classification. It classifies the severity according to the compromise of the area in question [26]. "Mild" is the compromise ≤1/3 of the normal area, "Moderate" is the compromise between 1/3 and 2/3 of the normal area, and "Severe" is the compromise >2/3 of the normal area. The interobserver agreement of the severity of the central and foraminal stenosis were substantial (k = 0.73) and moderate (k = 0.58), respectively. The interobserver agreement at the lateral recess stenosis had the worst agreement values, with an average kappa coefficient of 0.49 [26].

The severity of the stenosis, regardless of location, could be adequately assessed using the Lurie et al. classification. It classifies the severity based on the affected area as compared to the normal 26. "Mild" is the compromise ≤1/3 of the normal area, "moderate" is the compromise between 1/3 and 2/3 of the normal area, and "severe" is the compromise >2/3 of the normal area. The inter-observer agreement of the central and foraminal stenosis severity was substantial (k = 0.73) and moderate (k = 0.58), respectively. The inter-observer agreement at the lateral recess stenosis had the worst agreement values, with an average kappa coefficient of 0.49 [26].

Assessment of central and lateral recess stenosis severity is achieved through the dural sac morphology, employing classifications by Lee Guen et al. and Schizas et al. Lee Guen et al.'s classification comprising four grades (from grade 0 to grade 3), offers simplicity and exhibits superior intra- and interobserver agreement. Conversely, Schizas et al. introduced a more complex system with seven degrees. However, it has limitations; for instance, grade A combines patients without stenosis and those with mild stenosis. Moreover, grades C ('severe stenosis') and D ('extreme stenosis') have notably similar descriptions, differing primarily by the absence of posterior epidural fat. Another advantage of Lee Guen et al.'s classification lies in its correlation between severity degrees and the dural sac cross-sectional area. In 2013, Park et al. validated Lee Guen et al.'s classification and found an association between symptoms and the degree of stenosis [28].

For lateral recess stenosis, the four-grade classification system described by Pfirmann et al. [47] reported an inter-reader agreement of 0.62–0.67 among three readers including one spine surgeon and two radiologists. Kaliya-Perumal et al. [49], in 2018, did a revalidation of the grading system. They reported an inter-reader agreement of 0.521 among three orthopedic surgery residents. The modification proposed by Miskin et al. had an inter-observer agreement lower than these previously reported results (k=0.323) [48].

For us, the best classification for foraminal stenosis was proposed by Lee et al. [31]. Their MRI-based classification categorizes foraminal stenosis into four degrees based on morphology. In contrast, Wildermuth et al.'s classification30 offers less detailed descriptions for each degree compared to Lee et al.'s approach. Notably, Wildermuth's classification primarily focuses on changes in epidural fat, whereas Lee et al.'s system evaluates multiple factors, including epidural fat, stenosis type, and nerve compression presence [30,31]. Özer et al.'s classification [45] presents an intriguing approach, considering vertebral level stability and providing surgical guidance for each subgroup.

Degenerative Spondylolisthesis Classification Systems

We identified two imaging-based classification systems dedicated to degenerative spondylolisthesis. One, proposed by Gille et al. [34], derived from Schwab et al.'s "Adult Spinal Deformities Classification" [36]. The other, by Kepler et al. [33], known as the "CARDS classification", incorporates both clinical and radiological variables. Radiological parameters encompass disc space height, sagittal alignment, and disc translation. While clinical parameters consider pain in the lower extremities - unilateral or bilateral. Both classifications had good reliability studies [33,35,41], as outlined in Table 5. However, our judgment suggests their limitation lies in the complexity of routine clinical application. Kong et al. compared both classifications in a retrospective study. They concluded that both systems have acceptable reliability, but the CARDS classification was easier to use and had better inter and intra-rater agreement values. Their findings highlighted that type D in the CARDS classification and type 5 in Gille et al.'s system correlated with worse preoperative pain and showed greater post-surgery improvement. Notably, Gille et al.'s classification provides more comprehensive information for therapeutic decisions [41].

Conversely, Meyerding et al.'s classification [32] gains favor for its ease of application and widespread use due to its simplicity, supported by substantial intra and interobserver

Molina M and Vial S., J Spine Res Surg 2023
DOI:10.26502/fjsrs0068

agreement (k = 0.79 and 0.78, respectively) [42]. However, its primary limitation lies in the lack of consideration for morphological parameters such as segmental kyphosis or disc height, which bear significance in prognosis [43].

Based on the literature review, we recommend the following classifications due to their better intra and interobserver reproducibility and clinical application:

- **Lumbar Disc Degeneration Classifications Systems:** Pfirrmann et al. [7]
- **Lumbar Disk Herniation Classifications Systems:**
    - Location: Fardon et al. [39], MSU classification (Mysliwiec et al.) [15]
    - Morphology: Fardon et al. [39]
    - Degree: Lurie et al. [26], MSU classification (Mysliwiec et al.) [15]
- **Lumbar Facet Joint Osteoarthritis Classifications Systems:** Weishaupt et al. [19]
- **Degenerative Lumbar Spinal Stenosis Classifications Systems:**
    - **Location:** Fardon et al. [39]
    - **Central–lateral recess stenosis:** Lee Guen et al. [28], Lurie et al. [26]
    - **Lateral recess stenosis:** Lurie et al. [26]
    - **Foraminal stenosis:** Lee et al. [31]
- **Degenerative Spondylolisthesis Classification Systems:** CARDS et al. [33]

Limitations

We could not recommend some classification systems due to lack of reliability. Other studies are not focused on clinical outcomes and do not aid in treatment decisions.

We did not find any classifications that correlate with prognosis. Hence, it is left to the surgeon and his experience with which classification to use.

We only searched articles in the English language, which may limit the evaluation of an eventual good classification in a different language.

Conclusions

There are many classification systems, with advantages and disadvantages. Some of them were more widely used because they were easily applied and reliable.

For a classification to hold clinical value, it should exhibit high reliability, typically indicated by a Kappa or ICC value exceeding 0.60. Additionally, it should offer clinical guidance to aid in therapeutic decisions and integrate them into guidelines.

Our review revealed that existing classifications only partially provided the above characteristics. A combination of different classifications allows a better description of the pathology and may categorize patient's findings into subgroups that are similar in terms of prognosis and management. Further studies focused on classification development are needed to create improved systems with increased clinical utility and higher reliability.

Conflict of Interest

None of the authors has any potential conflict of interest.

Acknowledgment

This study was supported by the AO Spine Latin America. AO Spine is a clinical división of the AO Foundation, which is an independent medically guided Not-for-profit organization.

References

1. Cheung K.M.C, Karppinen J, Chan D, et al. Prevalence and Pattern of Lumbar Magnetic Resonance Imaging Changes in a Population Study of One Thousand Forty-Three Individuals. Spine 34 (2009): 934-940.
2. Kettler A, & Wilke H. Review of existing grading systems for cervical or lumbar disc and facet joint degeneration. European Spine Journal 15 (2005): 705-718.
3. Vaccaro A, Lehman R, Hurlbert R, et al. A New Classification of Thoracolumbar Injuries. Spine 30 (2005): 2325-2333.
4. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 33 (1977): 159-174.
5. Lane NE, Nevitt MC, Genant HK. Reliability of new indices of radiographic osteoarthritis of the hand and hip and lumbar disc degeneration. J Rheumatol 20 (1993): 1911-1918.
6. Schneiderman G, Flannigan B, Kingston S, et al(1987). Magnetic Resonance Imaging in the Diagnosis of Disc Degeneration. Spine 12 (1987): 276-281.
7. Pfirrmann CW, Metzdorf A, Zanetti M, et al. Magnetic resonance classification of lumbar intervertebral disc degeneration. Spine 26 (2001): 1873-1878.
8. Griffith J.F, Wang Y.X.J, Antonio G.E, et al. Modified Pfirrmann Grading System for Lumbar Intervertebral Disc Degeneration. Spine 32 (2007): E708-E712.
9. Riesenburger R.I, Safain M.G, Ogbuji R, et al. A novel classification system of lumbar disc degeneration. Journal of Clinical Neuroscience 22 (2015): 346-351.
10. Burke S.M, Hwang S.W, Mehan, W.A, et al. Reliability of the modified Tufts Lumbar Degenerative

fortune journals

Molina M and Vial S., J Spine Res Surg 2023
DOI:10.26502/fjsrs0068

Disc Classification between neurosurgeons and neuroradiologists. Journal of Clinical Neuroscience 29 (2016): 111-116.

11. Madan, Sanjeev & Rai, Am & Harley, et al. Interobserver Error in Interpretation of the Radiographs for Degeneration of the Lumbar Spine. The Iowa orthopaedic journal 23 (2003): 51-56.

12. Ahn Y, Jeong T.S, Lim T, et al. Grading system for migrated lumbar disc herniation on sagittal magnetic resonance imaging: an agreement study. Neuroradiology 60 (2017): 101-107.

13. Halldin K, Lind B, Rönnberg K, et al. Three-dimensional radiological classification of lumbar disc herniation in relation to surgical outcome. International Orthopaedics 33 (2008): 725-730.

14. Lee S, Kim S.K, Lee S.H, et al. Percutaneous endoscopic lumbar discectomy for migrated disc herniation: classification of disc migration and surgical approaches. 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 16 (2007): 431-437.

15. Mysliwiec L.W, Cholewicki J, Winkelpleck M.D, et al. MSU Classification for herniated lumbar discs on MRI: toward developing objective criteria for surgical selection. European Spine Journal, 19 (2010): 1087-1093.

16. Wiltse L.L, Berger P.E, McCulloch J.A. A System for Reporting the Size and Location of Lesions in the Spine. Spine 22 (1997): 1534-1537.

17. Hao D.J, Duan K, Liu T.J, et al. Development and clinical application of grading and classification criteria of lumbar disc herniation. Medicine, 96 (2017): e8676.

18. Pathria M, Sartoris DJ, Resnick D. Osteoarthritis of the facet joints: accuracy of oblique radiographic measurement. Radiology 164 (1987): 227-230.

19. Weishaupt D, Zanetti M, Boos N, et al. MR imaging and CT in osteoarthritis of the lumbar facet joints. Skeletal Radiol 28 (1999): 215-219.

20. Fujiwara A, Tamai K, Yamato M, et al. The relationship between facet joint osteoarthritis and disc degeneration of the lumbar spine: an MRI study. Eur Spine J 8 (1999): 396-401.

21. Little J.W, Grieve T.J, Cramer G.D, et al. Grading Osteoarthritic Changes of the Zygapophyseal Joints from Radiographs: A Reliability Study. Journal of manipulative and physiological therapeutics 38 (2015): 344-351.

22. Butler D, Trafimow J.H, Andersson G.B.J, et al. Discs Degenerate Before Facets. Spine 15 (1990): 111-113.

23. Coste J, Judet O, Barre O, et al. Inter- and intraobserver variability in the interpretation of computed tomography of the lumbar spine. Journal of Clinical Epidemiology 47 (1994): 375-381.

24. Grogan J, Nowicki B.H, Schmidt T.A, et al. Lumbar facet joint tropism does not accelerate degeneration of the facet joints. AJNR Am J Neuroradiol 18 (1997): 1325-1329.

25. Stieber J, Quirno M, Cunningham M, et al. The Reliability of Computed Tomography and Magnetic Resonance Imaging Grading of Lumbar Facet Arthropathy in Total Disc Replacement Patients. Spine 34 (2009): E833-E840.

26. Lurie J.D, Tosteson A.N, Tosteson T.D, et al. Reliability of Readings of Magnetic Resonance Imaging Features of Lumbar Spinal Stenosis. Spine, 33 (2008): 1605-1610.

27. Schizas C, Theumann N, Burn A, et al. Qualitative Grading of Severity of Lumbar Spinal Stenosis Based on the Morphology of the Dural Sac on Magnetic Resonance Images. Spine 35 (2010): 1919-1924.

28. Guen Y.L, Joon W.L, Hee S.C, et al. A new grading system of lumbar central canal stenosis on MRI: an easy and reliable method. Skeletal Radiology 40 (2011): 1033-1039.

29. Park, H.J, Kim S.S, Lee Y.J, et al. Clinical correlation of a new practical MRI method for assessing central lumbar spinal stenosis. The British Journal of Radiology 86 (2013): 20120180.

30. Wildermuth Zanetti M, Duewell S, Schmid M.R, et al. Lumbar spine: quantitative and qualitative assessment of positional (upright flexion and extension) MR imaging and myelography. Radiology 207 (1998): 391-398.

31. Lee S, Lee J.W, Yeom J.S, et al. A Practical MRI Grading System for Lumbar Foraminal Stenosis. American Journal of Roentgenology 194 (2010): 1095-1098.

32. Meyerding HW. Spondylolisthesis. Surg Gynecol Obstet 54 (1932): 371-377.

33. Kepler C.K, Hilibrand A.S, Sayadipour A, et al. Clinical and radiographic degenerative spondylolisthesis (CARDS) classification. The Spine Journal 15 (2015): 1804-1811.

34. Gille O, Challier V, Parent H, et al. Degenerative lumbar spondylolisthesis. Cohort of 670 patients, and proposal of a new classification. Orthopaedics & Traumatology: Surgery & Research 100 (2014): S311-S315.

35. Ghailane S, Bouloussa H, Challier V, et al. Radiographic Classification for Degenerative Spondylolisthesis of the Lumbar Spine Based on Sagittal Balance: A Reliability Study. Spine Deformity 6 (2018): 358-365.

Molina M and Vial S., J Spine Res Surg 2023
DOI:10.26502/fjsrs0068

36. Schwab F, Ungar B, Blondel B, et al. Scoliosis Research Society—Schwab Adult Spinal Deformity Classification. Spine, 37 (2012): 1077-1082.

37. Urrutia J, Besa P, Campos M, et al. The Pfirrmann classification of lumbar intervertebral disc degeneration: an independent inter- and intra-observer agreement assessment. European Spine Journal 25 (2016): 2728-2733.

38. Fardon D.F, Milette P.C. Nomenclature and Classification of Lumbar Disc Pathology. Spine 26 (2001): E93-E113.

39. Fardon D.F, Williams A.L, Dohring E.J, et al. Lumbar disc nomenclature: version 2.0. The Spine Journal 14 (2014): 2525-2545.

40. Berg L, Thoresen H, Neckelmann G, et al. Facet arthropathy evaluation: CT or MRI? European Radiology 29 (2019): 4990-4998.

41. Kong C, Sun X, Ding J, et al. Comparison of the French and CARDS classifications for lumbar degenerative spondylolisthesis: reliability and validity. BMC Musculoskeletal Disorders, 20 (2019).

42. Timon S.J, Gardner M.J, Wanich T, et al. Not All Spondylolisthesis Grading Instruments Are Reliable. Clinical Orthopaedics and Related Research 434 (2005): 157-162.

43. Sobol G.L, Hilibrand A, Davis A, et al. Reliability and Clinical Utility of the CARDS Classification for Degenerative Spondylolisthesis. Clinical Spine Surgery 31 (2018): E69-E73.

44. Zhu K, Su Q, Chen T, et al. Association between lumbar disc herniation and facet joint osteoarthritis. BMC Musculoskeletal Disorders 21 (2020).

45. Özer A.F, Akyoldaş G, Çevik O.M, et al. Lumbar Foraminal Stenosis Classification That Guides Surgical Treatment. International journal of spine surgery 16 (2022): 666-673.

46. Zhu F, Zhang Y, Peng Y, et al. A novel classification based on magnetic resonance imaging for individualized surgical strategies of lumbar disc herniation. Archives of orthopaedic and trauma surgery 143 (2023): 4833-4842.

47. Pfirrmann C.W, Dora C, Schmid M.R, et al. MR image-based grading of lumbar nerve root compromise due to disk herniation: reliability study with surgical correlation. Radiology 230 (2004): 583-588.

48. Miskin N, Isaac Z, Lu Y, et al. Simplified Universal Grading of Lumbar Spine MRI Degenerative Findings: Inter-Reader Agreement of Non-Radiologist Spine Experts. Pain medicine (Malden, Mass.) 22 (2021): 1485-1495.

49. Kaliya-Perumal A.K, Ariputhiran-Tamilselvam S.K, Luo C.A. Revalidating Pfirrmann's Magnetic Resonance Image-Based Grading of Lumbar Nerve Root Compromise by Calculating Reliability among Orthopaedic Residents. Clinics in orthopedic surgery 10 (2018): 210-215.

50. Fledelius, Joan & Khalil, Azza & Hjorthaug, et al. Inter-observer agreement improves with PERCIST 1.0 as opposed to qualitative evaluation in non-small cell lung cancer patients evaluated with F-18-FDG PET/CT early in the course of chemo-radiotherapy. EJNMMI Research 6 (2016).