CLINICAL-RADIOGRAPHIC CORRELATION OF DEGENERATIVE CHANGES OF THE SPINE – SYSTEMATIC REVIEW

ABSTRACT

Systematic review of the literature on the evaluation of images of degenerative changes of the spine and its clinical correlation. A systematic literature review was conducted, and the results evaluated for the presence of clinical correlation, as well as the type of imaging method used. The search terms were “Intervertebral Disc Degeneration”, “Intervertebral disc”, “Classification”, “Anulus fibrosus”, “Nucleus pulposus”, “Lumbar spine”, “Degenerative disc disease”, “Degeneration”, “Zygapophyseal Joint”. We also assessed whether there were inter- and intraobserver agreement in the selected works and possible guidelines regarding the treatment and prognosis of patients. Of the 91 reviewed abstracts, 31 articles were selected that met the inclusion criteria. Six articles were related to the cervical spine, 13 to the lumbar spine and 12 were about changes not related specifically to a single segment of the spine. Articles that determined limiting values considered normal were also included, since variations were considered signs of degeneration or pathology. It was not possible to establish the relationship between the changes identified in imaging and the clinical history of patients, either define treatment and prognosis guidelines.

Keywords: Intervertebral disc degeneration; Intervertebral disc; Classification; Spine; Zygapophyseal joint.

RESUMEN

Revisión sistemática de la literatura sobre evaluación por imagen de los cambios degenerativos de la columna y su correlación clínica. Se realizó una revisión sistemática de la literatura y sus resultados evaluados con relación a la presencia de correlación clínica, así como al tipo de método de imagen utilizado. Los términos de búsqueda fueron “Intervertebral Disc Degeneration”, “Intervertebral disc”, “Classification”, “Anulus fibrosus”, “Nucleus pulposus”, “Lumbar spine”, “Degenerative disc disease”, “Degeneration”, “Zygapophyseal Joint”. También se evaluó el acuerdo inter e intraobservador en los trabajos seleccionados y posibles orientaciones con relación al tratamiento y al pronóstico de los pacientes. De los 91 resúmenes revisados, 31 artículos que cumplieron los criterios de inclusión. Seis artículos estaban relacionados con la columna cervical, 13 con la columna lumbar y 12 describían cambios no relacionados específicamente con un único segmento de la columna. Los artículos que determinaron valores límite considerados normales también fueron incluidos, una vez que sus variaciones fueron consideradas señales de degeneración o patología. No fue posible establecer la relación entre los cambios identificados en los exámenes de imagen y la historia clínica de los pacientes, tampoco definir orientaciones de tratamiento y pronóstico.

Descritores: Degeneración del disco intervertebral; Disco intervertebral; Clasificación; Columna vertebral; Articulación zigapofisaria.

RESUMEN

Revisión sistemática de la literatura sobre evaluación por imágenes de los cambios degenerativos de la columna y su correlación clínica. Se realizó una revisión sistemática de la literatura y sus resultados evaluados con relación a la presencia de correlación clínica, así como al tipo de método de imagen utilizado. Los términos de búsqueda fueron “Intervertebral Disc Degeneration”, “Intervertebral disc”, “Classification”, “Anulus fibrosus”, “Nucleus pulposus”, “Lumbar spine”, “Degenerative disc disease”, “Degeneration”, “Zygapophyseal Joint”. Se evaluó también la concordancia inter e intraobservador en los trabajos seleccionados, además de las posibles orientaciones en cuanto al tratamiento y al pronóstico de los pacientes. De los 91 resúmenes revisados, treinta y un artículos que cumplieron los criterios de inclusión fueron seleccionados. Seis artículos estaban relacionados con la columna cervical, 13 con la columna lumbar y 12 describían cambios no relacionados específicamente a un solo segmento de la columna vertebral. También se incluyeron artículos que determinaron valores límite considerados normales, y que los cambios fueron considerados señales de degeneración o patología. No fue posible establecer la relación entre los cambios identificados en las imágenes y la historia clínica de los pacientes, tampoco definir orientaciones de tratamiento y pronóstico.

Descritores: Degeneración del disco intervertebral; Disco intervertebral; Clasificación; Columna vertebral; Articulación zigapofisaria.
INTRODUCTION

Degenerative changes in the spine are often found in imaging exams, independently of morphological and clinical changes among individuals of the same sex, age, and profession. The principal morphological changes observed in the intervertebral discs and facet joints have been the subject of various descriptions using histological1-9 and morphological10-15 methods and especially with the use of imaging methods, such as radiography (X-ray),10,13-15 computed tomography (CT),6,15-17 magnetic resonance imaging (MRI),6,15,18 and discography.19

The correlation between degenerative changes and the presence of symptoms is the subject of discussion and controversy. The prevalence of lumbar and cervical pain in the adult population and the ease of access to imaging exams create an often unfounded impression of disease. However, we lack the tools to enable us to distinguish the natural changes of the aging process from changes associated with pain and disability.

The need to classify the degeneration by grades is due to the fact that these systems can clarify doubts about the relationship between morphology and functional parameters, or between instability and intradiscal pressure, or be used to correlate the different morphological changes encountered.20

Several descriptions of disc degeneration are available in the literature, however these changes need to be quantified, especially for research reasons.21 An analysis of existing systems reveals different methods for grading the changes and possible correlations between lesions and symptoms, though there is currently no consensus.

Thus, the objective of the authors is to conduct a systematic review of the pertinent literature in search of objective evaluations of disc and facet joint degeneration, comparing their advantages and disadvantages, as well as highlighting their possible clinical correlations. We also intend to define a possible application of these evaluations to the decision-making process for the treatment of degenerative spine changes.

METHOD

A bibliographical investigation of the Medline, Embase, FirstBook, Web of Science, Google Scholar data bases and the Cochrane systematic reviews was performed using the search terms “Intervertebral Disc Degeneration”, “Intervertebral disc”, “Classification”, “Annulus fibrosus”, “Nucleus pulposus”, “Lumbar spine”, “Degenerative disc disease”, “Degeneration”, and “Zygapophyseal Joint”. Articles published during the period between 1945 and 2012 in English, French, German, Italian, Portuguese, and Spanish were included. We selected abstracts that present a description of degenerative spinal changes identified and quantified by imaging methods, in addition to studies that associate these changes with some severity scoring or grading system.

For the initial selection, the exclusion criteria applied were the detection of subjective and undetailed descriptions, metric units or scores that did not permit the degenerative changes to be quantified, articles in a language other than those cited above, and publications without an abstract available for consultation.

All the abstracts were first evaluated by three independent reviewers and, once compliance with the criteria was confirmed, complete versions of the selected articles were obtained. A critical reading of these followed and their respective references were researched for additional data, with the intention of refining the initial selection. Any disagreements about the selection of the articles were resolved by consensus among the three reviewers. All the articles used in this systematic review were categorized by level of evidence according to the scale of evidence of Sackett et al.22 This study is a systematic review of the literature and does not involve patients, so the approval of the Institutional Review Board was not required.

RESULTS

In the initial search, we found 681 articles and selected 92 of them using the abstracts. Of these, 31 satisfied the inclusion criteria. Of the 61 works excluded, four had no available abstract, eight were in languages other than those previously cited, and 49 lacked a quantified and objective description of vertebral degeneration.

Of the 31 articles selected, six included cervical measurements (four with values considered normal and two with degenerative changes), 13 lumbar measurements (eight indicating degenerative changes and five describing normal values), and 12 with changes not specifically related to a defined spinal segment, such as measurements related to the sacro-pelvic region, which can directly impact degeneration of spinal segments. (Figure 1)

The four articles with descriptions of cervical measurements23-26 identified normal values for parameters such as angulations and amplitudes of movement. These values were correlated with degenerative conditions, situations that lead to possible instability, pain, and neurological changes. (Table 1)

Three articles27-29 defined the normal values for the sacro-pelvic region, where a correlation with sagittal imbalance, the presence of a propensity towards spondylolisthesis and its complications (such as difficulty walking), neurological changes, and pain were demonstrated. (Table 2)

Roussouly et al.30 grouped the most common changes in sagittal alignment, the sacrum, and the pelvis into four groups, suggesting that for each group there would be a different disposition towards degeneration of the lumbar spine, whether facetectomy, discal, or a predisposition towards spondylolisthesis. (Figure 2) The authors also defined a subgroup that was considered normal and with less propensity towards degeneration.

Among all the articles reviewed, a description and respective assessment of the degenerative changes in the radiographs was found in six of them.27-31 It is worth noting the use of the same parameters for degeneration – the presence of osteophytes, the loss...
of disc height, and subchondral sclerosis - qualified in such a way as to be used for the composition of grading schemes regarding the severity of the clinical profile presented by the patient. (Table 3)

Pathria et al.36 developed a grading system based on degenerative facet changes identified in oblique radiographs and CT. Weishaupt et al.37 correlated changes in CT with those found in MRI, using the parameters of facet joint space, articular hypertrophy, and the presence of osteophytes to define their grades. Using MRI, Stadnik et al.38 identified degenerative changes, describing the concept as High Intensity Zones (HIZ), positively correlated with disc rupture.

Three other studies proposed grading systems using changes found in MRI,39–41 with the methodological particularities between them. Modic et al.39 evaluated changes in the vertebral plateau, in both T1- and T2-weighted sequences and Miyazaki et al.41 also addressed degenerative cervical changes in T2. Adams et al.42 in turn, described a classification of degeneration using image patterns found following discography exams, outlining the changes encountered and correlating them with the different evolutionary stages of disc degeneration. (Table 4).

In terms of the grading systems found in the review, (Table 5) we confirmed a disparity in the number of evolutionary stages of disc generation used, ranging from three to five. Additionally, the evaluation methodologies applied were not standardized, given that some studies used the presence or absence of changes as a scoring criterion, while others attributed values to them based mainly on their size and location. In addition, some of the systems proposed started their grading scales with the lowest intensity encountered, while others applied the inverse, with the most serious stage first.

The final results published also presented differences in the definitions of “the absence of changes”, being categorized as Grade 0 in some articles and as Grade 1 in others. It should be noted that studies using subjective terms like “mild”, “moderate”, and “initial”, among others, without specific criteria based on measurements or standardized images, were excluded from the results presented because they did not permit inter- and intraobserver reproducibility. No studies exclusively describing thoracic spine changes were found.

**DISCUSSION**

The large number of studies found in the literature demonstrates the need to identify the degenerative changes found in the spine, quantify them, and correlate them with the clinical profile, the treatment, and the prognosis of the patients. However, it also demonstrates the lack of a precise evaluation method that is directly linked to the clinical profile of the patient. Many times, changes considered to be advanced present less important symptoms in certain patients as compared to others with fewer changes in the image exams.

In another attempt at standardization based on clinical correlation, Modic et al.42 identified signs of discal fissure (HIZ) in 56% of the asymptomatic individuals in MRI with and without contrast and this sign was considered to be a weak predictor of the occurrence of lumbar pain.

A grading system needs to meet several basic criteria to be considered valid. The score needs to be the same regardless of the examiner’s experience and must reflect the real grade of degeneration.21 Most of the studies base their results on subjective evaluations, using criteria such as “moderate” or “severe” or “small” or “large”13,34,43-45.

Most of the studies adopt an evaluation of the changes through radiographic exams. Wilke et al.21 propose three advantages in relation to this methodology: they are less invasive than other exams; they
are simpler; they are less expensive and often available because of previous diagnostic studies or for patient follow-up.

Even with the heightened interest in this topic, we confirmed that there is a lack of consensus around the criteria evaluated in degenerative changes. Most authors verified radiographic exams in anteroposterior (AP) and lateral (L) views, to look for the presence of osteophytes and subchondral sclerosis, as well as to assess disc height and the dimensions and positions of the osteophytes.

Frobin et al.45 collaborated in an important way by defining normal disc height values, later being able to apply them as a populational reference. For this reason, in our study the patients were categorized into only three grades of degeneration – normal, mild, and moderate/severe. In an optimization process, Mimura et al.33 (Table 6) proposed a new system, also based on spinal radiographs. For this, they evaluated disc height (divided into four different anatomical sections), the presence of osteophytes and their size, and the presence of vertebral sclerosis in one or both plateaus. The main objective of the grading system was to investigate the effects of degeneration on the flexibility of the spine, but there is no description of clinical tests to validate it.

Wilke et al.21 described a similar grading system for lumbar changes, (Table 7) in which the criteria were scored individually and the final grade of degeneration was determined by summing these three items. Interobserver agreement, which according to the Kappa coefficient was 0.714 points indicating substantial agreement, was tested among different evaluators. The authors concluded the study, calling the proposed system “almost” objective, valid, and reliable, but they noted that the true grade of degeneration tends to be underestimated, and that there are significant differences in evaluations depending on the experience of the evaluator.

Lane et al.34 (Table 8) evaluated radiographs in lateral view only, addressing the narrowing of the articular space, the presence of osteophytes, and sclerosis of the vertebral plateaus. The interobserver validity analysis, based on the assessments of three evaluators considered to be experienced, yielded Kappa coefficients of 0.93 for reduction of the articular space, 0.91 for the osteophytes, and 0.93 for the final result. However, in relation to sclerosis of the vertebral plateau, a Kappa coefficient of only 0.55 was obtained, indicating weak agreement between the evaluators for the item in question.

Kettler et al.35 proposed a new grading system for cervical degeneration, similar to the one developed for the lumbar spine by Wilke et al.31 They evaluated the loss of disc height and the presence of osteophytes and subchondral sclerosis, and they compared the radiographic scores with the structural changes found macroscopically after dissecting the pieces. The value of the Kappa coefficient found was 0.688. Among all the criteria evaluated, the single variable with the highest agreement was the presence of osteophytes and subchondral sclerosis, as well as to assess disc height respecting the individual height prior to degeneration.

This proposed method of analysis, the values of the Kappa coefficient ranged from 0.74 to 0.81. Theses higher indices, as compared to those obtained from the radiographic classifications, suggest that the items evaluated via MRI and the intensity of the signal may be more objective than the bone changes found in radiographs.

Regarding the reliability of the validation studies of the proposed grading systems, Kovacs et al.30 suggested that, in various cases, high agreement is due to the reduced number of observers and to the fact that they belong to the same hospital service. In their study, seven observers at different levels of experience in spine surgery from three distinct services evaluated the imaging exams of the spine applying the grading proposed by Pfirrmann and found an intraobserver agreement of 0.601 (Kappa ranging from 0.00 to 0.99) and an interobserver agreement of 0.219 (Kappa ranging from 0.024 to 0.59), demonstrating little interobserver agreement for this classification.

In the same study, Kovacs et al.30 also assessed the agreement between the Modic grading system and the presence of the HIZ signal in MRI. High inter- and intraobserver agreements were identified; however, the literature shows that the presence of these changes is a weak predictor of future low back pain, conferring little prognostic value. Today we

### Table 6. Grading system of Mimura et al.

<table>
<thead>
<tr>
<th>Loss of disc height (% of adjacent disc)</th>
<th>Osteophytic formation*</th>
<th>Sclerosis of the vertebral plateaus</th>
<th>Results (sum of points)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 = Normal</td>
<td>0 = 0 Points</td>
<td>0 = Absent</td>
<td>Grade I: 0 to 1.5</td>
</tr>
<tr>
<td>1 = Mild (&gt;75%)</td>
<td>1 = 1-4 Points</td>
<td>1 = One plateau (above or below)</td>
<td>Grade II: 2.0 -3.0</td>
</tr>
<tr>
<td>2 = Moderate (&gt;50%)</td>
<td>2 = 5-8 Points</td>
<td>2 = Both plateaus</td>
<td>Grade III: 4.0-6.0</td>
</tr>
<tr>
<td>3 = Severe (&gt;25%)</td>
<td>3 = 9-12 Points</td>
<td></td>
<td>Grade IV: &gt; 6.0</td>
</tr>
<tr>
<td>4 = Very Severe (&lt;25%)</td>
<td>4 = 13-16 Points</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* = sum of the points of the eight extremities; <3mm = 1 point; >3mm = 2 points.

### Table 7. Lumbar spine grading system of Wilke et al.21

<table>
<thead>
<tr>
<th>Loss of height</th>
<th>Osteophytic formation</th>
<th>Diffuse sclerosis</th>
<th>Total grade of degeneration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sum of the points of the 2 extremities</td>
<td>Sum of the points of the adjacent bodies</td>
<td>No sclerosis: 0 points</td>
<td>Sum of the points from “Loss of height,” “Osteophyte Formation,” and “Diffuse Sclerosis.”</td>
</tr>
<tr>
<td>&lt;3 mm = 1 point</td>
<td>0.25 partially or completely affected: 1 point</td>
<td>0.5 partially or completely affected: 2 points</td>
<td></td>
</tr>
<tr>
<td>≤3 and &lt; 6 mm = 2 points</td>
<td>≤6 = 3 points</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;0.5 partially or completely affected: 3 points</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 8. Grading system of Lane et al.34

<table>
<thead>
<tr>
<th>Narrowing of articular space</th>
<th>Anterior and posterior osteophytes</th>
<th>Sclerosis of the vertebral plateau</th>
<th>Grading of the degeneration</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 = None</td>
<td>0 = None</td>
<td>0 = None</td>
<td>Grade 0 = Normal articulation (0 for osteophytes and narrowing)</td>
</tr>
<tr>
<td>1 = Mild</td>
<td>1 = Small</td>
<td>1 = Present</td>
<td>Grade I = Moderate (1) narrowing or mild (1) osteophytes</td>
</tr>
<tr>
<td>2 = Moderate</td>
<td>2 = Moderate</td>
<td></td>
<td>Grade II = Moderate- Severe (2-3) narrowing and/or moderate to severe (2-3) osteophytes</td>
</tr>
</tbody>
</table>
know that HIZ can be present in the imaging exams of 30% to 56% of asymptomatic individuals.37,38 Considering the studies evaluated, it is possible that, although the criteria for the definition of disc degeneration have been established, this specificity may differ from the visual impression regarding the ratio of disc height at different times of day and at different patient ages, since significant variations have been described even during a single day.45-47 Other items evaluated, such as the presence and size of the osteophyte, while lending a higher degree of objectivity to the evaluation, reflect significant variations among observers in the literature.

CONCLUSION

Several grading systems involving geometric or descriptive measurements of changes related to disc and facet degeneration are available in the literature, however, there is great disparity in terms of the criteria applied, the results presented, and the possible clinical correlations. No single system was identified in this review that can define prognosis or guide treatments. It was also confirmed that most of the studies proposed have not been tested for their validity or inter- and intraobserver agreement. A significant disparity remains in the evaluation methodologies, both for the most efficient imaging exam, taking cost versus benefit into account, and for the ideal criteria to be evaluated.

Thus, to date, according to the available literature, future research is needed to define the ideal system. All the authors declare that there are no potential conflicts of interest regarding this article.

CONTRIBUTIONS OF THE AUTHORS: Each author made significant individual contributions to the development of the manuscript. ENV participated in the development of the research project, the review of the literature, and the preparation of the manuscript. LRGV participated in the review of the manuscript. CES and PSM participated in the literature review and the preparation of the manuscript.

REFERENCES


