Classification of intervertebral disk degeneration using quantitative T2 relaxation time measurements

Poster No.: C-1685  
Congress: ECR 2013  
Type: Scientific Exhibit  
Authors: S. A. Nagy¹, I. Juhasz¹, H. Komaromy¹, K. Pozsar², I. Zsigmond², M. Aradi³, P. Gabor¹, O. Gergely¹, P. Bogner¹; ¹Pecs/HU, ²Budapest/HU, ³Zalaegerszeg/HU  
Keywords: Biological effects, Molecular imaging, MR, Image manipulation / Reconstruction, Neuroradiology spine  
DOI: 10.1594/ecr2013/C-1685

Any information contained in this pdf file is automatically generated from digital material submitted to EPOS by third parties in the form of scientific presentations. References to any names, marks, products, or services of third parties or hypertext links to third-party sites or information are provided solely as a convenience to you and do not in any way constitute or imply ECR’s endorsement, sponsorship or recommendation of the third party, information, product or service. ECR is not responsible for the content of these pages and does not make any representations regarding the content or accuracy of material in this file.

As per copyright regulations, any unauthorised use of the material or parts thereof as well as commercial reproduction or multiple distribution by any traditional or electronically based reproduction/publication method is strictly prohibited.

You agree to defend, indemnify, and hold ECR harmless from and against any and all claims, damages, costs, and expenses, including attorneys’ fees, arising from or related to your use of these pages.

Please note: Links to movies, ppt slideshows and any other multimedia files are not available in the pdf version of presentations.

www.myESR.org
**Purpose**

Several MRI grading systems have been used to characterize intervertebral disc degeneration (IVDD) in humans. Most classification systems for IVDD focus on signal intensity and morphology of the nucleus pulposus on sagittal T2-weighted MRI [1-6], because decreased signal intensity correlates with the degree of degenerative process, including the loss of water and proteoglycan content. Recently, the most widely accepted grading system was described by Pfirrmann et al. which is based on the distinction between the signal intensity of nucleus pulposus and annulus fibrosus, disc structure and disc height [4].

Nevertheless, morphology-based grading of IVDD suffers from two substantial handicaps. On one hand, T2 relaxation process might be affected by several molecular events as a result of subtle changes in the nucleus pulposus. On the other, the visual perception of signal intensity changes on T2-weighted images is problematic, because many arbitrary factors in signal detection and amplification may lead significant differences between raters causing inter-observer bias, especially when observers classify minor changes. The alterations between observers in terms of rating behavior, mostly due to the consistent over- or under estimation of the grading boundaries.

Based on the biophysical and biochemical changes in discs, several MR imaging techniques have been developed for quantitative evaluation of IVDD such as magnetization transfer imaging [7, 8], T1 and T1rho imaging [9-11], spectroscopy [12], diffusion weighted imaging [13, 14], and T2 measurements [15, 16]. Among these methods, T2 relaxation time evaluation provides the most reliable correlation with changes of molecular environment in the disc. The increase in water or proteoglycan content in the disc is associated with increased T2 values, while the increase in collagen content gives decreased T2 relaxation time [17]. Despite of the availability of different quantitative methods, in most studies IVDD is still graded with the semiquantitative morphological method of Pfirrmann et al, but other grading systems are also used. The grading in these studies is based usually on the evaluation of one or two raters and one morphological classification system [17-19]. As yet, only a study with limited scope by Takashima et al has suggested that T2 mapping may have a role to determine the range of T2 values for disc degeneration [20].

Since the degenerative pathology of intervertebral disc is related to low back pain [21, 22], the development of a classification system based on quantitative assessment would help to detect the degeneration level. Thus, the aim of this study was to investigate the possible inter-observer differences and to define quantitative T2 cut-off values with regard to morphological classification systems in patients with intervertebral disc degeneration.
Methods and Materials

Subjects

Twenty-one patients (13 female, 8 male; age range: 14-80 years; mean age 51.14±18.88 years) with single or recurrent episode of low back pain were examined. All the patients included in this study presented initially to the outpatient spine clinic, and then were referred for a lumbar MRI. Exclusion criteria were contraindication for MRI and previous malignant disease or injury in the lumbar spine. Inclusion criteria were no spine surgery in the last 10 years. No patients had evidence of lumbar disorders except intervertebral disc degeneration or herniation. One disc (L5-S1) of a patient was excluded from further analysis due to previous back surgery in the last 6 month. For the evaluation of age-related changes, all patients were dichotomized into two subgroups based on the median age of 51 years, as a younger group (6F/5M; age range: 14-51 years; mean age 37.2±14.3 years) and an older group (7F/5M; age range: 55-80 years; mean age 66.5±8.1 years).

Magnetic resonance imaging

The measurements were carried out using a 3T Magnetom TIM Trio whole-body MRI scanner (Siemens AG, Erlangen, Germany) with a gradient strength of 45 mT/m using a dedicated eight-channel spine coil (3T Spine matrix Coil, Siemens AG, Erlangen, Germany). Both morphological and quantitative MRI was performed covering the intervertebral discs L1-L2 to L5-S1. The imaging protocol included sagittal T2-weighted turbo spin echo (repetition time [TR] 5000 msec; echo time [TE] 105 msec; slice thickness 4 mm; interslice gap 0.4 mm; field of view [FOV] 280x280; matrix 269x384; bandwidth [BW] 260 Hz/px; number of excitations [NEX] 2; echo train length [ETL] 31) and T1-weighted turbo spin echo images (TR 640 msec; TE 11 msec; slice thickness 4 mm; interslice gap 0.4 mm; FOV 217x280; matrix 223x320; BW 252 Hz/px; NEX 4; ETL 6). Axial T2-weighted turbo spin echo images (TR 6123; TE 93; slice thickness 4 mm; interslice gap 0.4; FOV 171x220; matrix 196x320; BW 220 Hz/px; NEX 3; ETL 18) and T1-weighted turbo spin echo images (TR 797 msec; TE 12 msec; slice thickness 4 mm; interslice gap 0.4 mm; FOV 171x220; matrix 196x320; BW 279 Hz/px; NEX 2; ETL 10) were also performed. All the sequences were acquired without fat suppression.

Quantitative multi-echo T2 analysis was performed using a Carr Purcell Meiboom Gill imaging sequence in the sagittal section with 180° slice selective pulses and full phase rewinding in between the refocusing pulses: TR 2370 msec; echo spacing 10 msec; ETL 20; slice thickness 6 mm; interslice gap 1 mm; FOV 256x256; matrix 256x256; BW 186
Hz/px; NEX 1. The images were analyzed in the sagittal direction with the voxel size of 1 x 1 x 6 mm.

Data analysis

Morphological evaluation

All discs were classified morphologically on sagittal TSE T2-weighted images by three experienced neuroradiologists according to Pfirrmann, Schneiderman classification systems and other scheme based on Pfirrmann and Southern changes for degenerative intervertebral discs (Pfirrmann type classification). The three neuroradiologists blindly classified the grade of disc degeneration as grade I to V in the midsagittal section according to Pfirrmann classification [4] (Table 1), while Pfirrmann grade IV was subdivide into IVa and IVb in Pfirrmann type classification [23]. Schneiderman category was classified as grade 0 to 3 depending on the degeneration level (Grade 0, normal height and signal intensity; Grade 1, speckled pattern or heterogeneous decreased intensity; Grade 2, diffuse loss of signal; and Grade 3, signal void) [1]. The morphological classification of disc degeneration was assessed independently (differences were not settled by consensus) on a preselected echo where T2 weighting was prominent. The evaluation of inter-observer differences was performed by three neuroradiologists with 15, 18, 21 years of experience, respectively. To facilitate the calculation of inter-observer agreement, a total IVDD score was calculated for each individual as a rounded average of the five discs.

Quantitative assessment

For T2 measurements, region of interests (ROIs) were drawn over the whole area of nucleus pulposus on an echo where the distinction between nucleus pulposus (NP) and anulus fibrosus was clearly defined (Fig. 1). To minimize the errors, all measurements were performed by a single investigator (IJ). The ROIs were sampled on a given slice of the midsagittal section of NP to avoid possible partial volume effects. As a result of multi-echo spin echo (CPMG) sequence, the image consist of 20 single echo images with TE ranging from 10 msec to 200 msec in 10 msec intervals. T2 relaxation time data processing was carried out in each ROI with non-negative least squares algorithm (NNLS) using Matlab® software's curve fitting toolbox and a self-written program code (The MathWorks, Inc., Natick, MA). For T2 measurement, the first and every second echoes was excluded to minimize error from stimulated echoes [24]. The signal at each echo was measured and plotted as a function of time on a semi-log scale where no curvature was found; therefore T2 relaxation time was measured by calculating and mono-exponentially fitting the mean signal intensity using the following equation:
\[ S_i(t) = S_0 \exp\left(-\frac{TE}{T2}\right), \]

where \( i \) is used to denote the \( i^{th} \) image with echo time \( TE_i \) (\( i=2, 4, 6 \ldots 10 \)) and \( S_0 \) is proportional to the proton density.

**Statistical analysis**

Data analysis was performed using SPSS statistical software®, version 19.0 (SPSS, Inc., Chicago, IL). Multiple linear regression model was used to determine whether IVDD and T2 values were related to such predictors as age, gender, herniation and previous law back surgery. Predictors were introduce in a forward stepwise manner and retained in the model only if \( p \leq 0.05 \) was achieved. The assumptions of multiple linear regressions were satisfied, as judged by testing for linearity, normality assumptions of the residues, outliers, independence of errors, homoscedasticity and multi-collinearity [25]. Inter-observer agreement among multiple raters was calculated according to generalized Kappa statistics based on the standard least square approach [26], while Chi-square analysis was performed to determine whether the observed counts significantly differ from the expected counts in the examined grading systems. Generalized kappa statistics were interpreted as follows: values between 0 and 0.2 indicate slight agreement, 0.21-0.4 indicates fair agreement, 0.41-0.6 indicates moderate agreement, 0.61-0.8 indicates substantial agreement and >0.81 indicates almost perfect agreement [27]. In addition, receiver operating characteristic (ROC) analysis was performed among grades to determine the T2 cut-off values in each classification. A level of \( p \leq 0.05 \) was the significance threshold for all statistical analysis.

**Images for this section:**

<table>
<thead>
<tr>
<th>Grade</th>
<th>Structure</th>
<th>Distinction of the nucleus and annulus</th>
<th>Signal intensity</th>
<th>Height of the intervertebral disk</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Homogeneous, bright white</td>
<td>Clear</td>
<td>Hyperintense, isointense to CSF</td>
<td>Normal</td>
</tr>
<tr>
<td>II</td>
<td>Heterogeneous with or without horizontal bands</td>
<td>Clear</td>
<td>Hyperintense, isointense to CSF</td>
<td>Normal</td>
</tr>
<tr>
<td>III</td>
<td>Heterogeneous, gray</td>
<td>Unclear</td>
<td>Intermediate</td>
<td>Normal to slightly decreased</td>
</tr>
<tr>
<td>IV</td>
<td>Heterogeneous, gray or black</td>
<td>Lost</td>
<td>Intermediate or hypointense</td>
<td>Normal to moderately decreased</td>
</tr>
<tr>
<td>V</td>
<td>Heterogeneous, black</td>
<td>Lost</td>
<td>Hypointense</td>
<td>Collapsed disk space</td>
</tr>
</tbody>
</table>

**Table 1:** Intervertabral disc degeneration according to Pfirrmann et al. [4]
Fig. 1: Region of interest localization in the nucleus pulposus (TE=190 msec)
Results

In total, 21 patients with IVDD were evaluate using sagittal sections of T2 weighted MRI images by 3 independent neuroradiologists. Age dependent changes, inter-observer agreement could be calculated at a total of 104 intervertebral discs. Among the 105 discs, one disc had to be excluded from further post processing due to previous back surgery in the preceding 6 month.

Morphological assessment of age and grading

According to Pfirrmann, Schneiderman and Pfirrmann type classifications the average level of IVDD was significantly higher (p<0.001) in older patients compared with younger group. The grades of IVDD significantly increased as a function of age for all morphological categories (Fig.2 A, B and C).

Inter-observer agreement

Significant differences were found between the observers for the grading measurements in Pfirrmann category according to Chi-square analysis (Cramer's V value = 0.172; p=0.018; agreement 49%). In total, the highest number of cases was measured in Pfirrmann 3 according to the second and third observer (Fig. 3A). In Schneiderman classification no differences were found between the observers. Pfirrmann type classification showed a highly significant difference between the observers (Cramer's V value = 0.258 p <0.0001; agreement 43%). Figure 3B shows the distribution of the counts. For evaluation of individual grades, the lowest inter-observer agreement was found in Pfirrmann grades III and IV, while IVa and IVb showed the highest disagreement in Pfirrmann type classification (Table 2).

Table 2. Multi-rater generalized kappa values with upper and lower 95% confidence intervals.

<table>
<thead>
<tr>
<th>Classifications</th>
<th>Grades</th>
<th>Generalized kappa ±ASE</th>
<th>95% Confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pfirrmann</td>
<td>overall</td>
<td>0.28±0.08</td>
<td>0.13-0.43</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>0.48±0.42</td>
<td>-0.35-1.31</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>0.42±0.31</td>
<td>-0.19-1.03</td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>0.21±0.36</td>
<td>-0.49-0.92</td>
</tr>
</tbody>
</table>
Quantitative T2 analysis: age and morphological classifications

The mean T2 value in the 104 lumbar discs of IVDD was 78.35±24.20 msec. Multiple linear regression analysis revealed a highly significant inverse correlation between T2 values and age (Fig. 4).

T2 values of disc degeneration were as follows according to Pfirrmann classification: grade I (>119.8 msec), grade II (90.16-119.7 msec), grade III (64.5-90.16 msec), and grade IV (58.61-64.50 msec). Here, T2 times showed a significant difference between grade III and IV (<64.50 msec; sensitivity 100%, specificity 75%), while the resulting kappa coefficients were only 0.21 and 0.19, respectively. Despite the fair inter-observer agreement in Schneiderman grade 1 and 2, T2 values revealed the differences with 100% sensitivity and 100% specificity. T2 cut-off value showed the quantitative range of the degeneration level in Pfirrmann type IVa (58.15-75.37 msec) and IVb (49.02-58.15 msec), while the observers slightly agreed with each other (Table 3).

Table 3. Receiver operating characteristic analysis of T2 cut-off values with regard to morphological classifications.

<table>
<thead>
<tr>
<th></th>
<th>IV</th>
<th>Pfirrmann type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>V</td>
<td>overall</td>
</tr>
<tr>
<td>IV</td>
<td>0.19±0.32</td>
<td>0.04±0.32</td>
</tr>
<tr>
<td>V</td>
<td>0.34±0.33</td>
<td>0.28±0.06</td>
</tr>
<tr>
<td>Schneiderman</td>
<td>overall</td>
<td>0.40±0.08</td>
</tr>
<tr>
<td>0</td>
<td>0.73±0.35</td>
<td>0.48±0.42</td>
</tr>
<tr>
<td>1</td>
<td>0.30±0.32</td>
<td>0.61±0.31</td>
</tr>
<tr>
<td>2</td>
<td>0.22±0.37</td>
<td>0.30±0.31</td>
</tr>
<tr>
<td>3</td>
<td>0.63±0.31</td>
<td>0.20±0.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>overall</td>
</tr>
<tr>
<td>I</td>
<td>0.48±0.42</td>
<td>0.17±0.31</td>
</tr>
<tr>
<td>II</td>
<td>0.61±0.31</td>
<td>0.20±0.35</td>
</tr>
<tr>
<td>III</td>
<td>0.30±0.31</td>
<td>0.20±0.35</td>
</tr>
<tr>
<td>IVa</td>
<td>0.04±0.32</td>
<td>0.04±0.32</td>
</tr>
<tr>
<td>IVb</td>
<td>0.17±0.31</td>
<td>0.17±0.31</td>
</tr>
<tr>
<td>V</td>
<td>0.20±0.35</td>
<td>0.20±0.35</td>
</tr>
</tbody>
</table>

Note: ASE; asymptotic standard error

*Inter-observer reliability is displayed between only the first readings of all observers.
### Categories Grades

<table>
<thead>
<tr>
<th>Grading System</th>
<th>Pair</th>
<th>T2values</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>AUC</th>
<th>p-values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pfirrmann</td>
<td>I vs II</td>
<td>&lt;119.8</td>
<td>100</td>
<td>100</td>
<td>1</td>
<td>&lt;0.005</td>
</tr>
<tr>
<td></td>
<td>II vs III</td>
<td>&lt;90.16</td>
<td>100</td>
<td>100</td>
<td>1</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>III vs IV</td>
<td>&lt;64.50</td>
<td>100</td>
<td>75</td>
<td>0.875</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td></td>
<td>IV vs V</td>
<td>&lt;58.61</td>
<td>72.73</td>
<td>75</td>
<td>0.7273</td>
<td>ns</td>
</tr>
<tr>
<td>Schneiderman</td>
<td>0 vs 1</td>
<td>&lt;110.3</td>
<td>90.91</td>
<td>100</td>
<td>0.9899</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>1 vs 2</td>
<td>&lt;86.30</td>
<td>100</td>
<td>100</td>
<td>1</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>2 vs 3</td>
<td>&lt;53.79</td>
<td>68.75</td>
<td>75</td>
<td>0.7813</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Pfirrmann type</td>
<td>I vs II</td>
<td>&lt;119.8</td>
<td>100</td>
<td>100</td>
<td>1</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>II vs III</td>
<td>&lt;90.16</td>
<td>100</td>
<td>100</td>
<td>1</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td></td>
<td>III vs IVa</td>
<td>&lt;75.37</td>
<td>74.29</td>
<td>60</td>
<td>0.6343</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>IV vs IVb</td>
<td>&lt;58.15</td>
<td>69.23</td>
<td>71.43</td>
<td>0.7821</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>IVb vs V</td>
<td>&lt;49.02</td>
<td>77.78</td>
<td>76.92</td>
<td>0.741</td>
<td>&lt; 0.05</td>
</tr>
</tbody>
</table>

Note: ROC analysis of T2 values was performed among grades when the decision of the observers was in "consensus" with each other.

AUC; area under the curve

**Images for this section:**

![Fig. 2](image-url)

**Fig. 2:** Grading scores as an average of three independent observers revealed a significant elevation of degenerative changes with age according to Pfirrmann (A), Schneiderman (B) and Pfirrmann type classifications (C). All the significance levels are adjusted for gender, herniation and previous low back surgery.
Fig. 3
**Fig. 4:** All the significance levels are adjusted for gender, herniation and previous low back surgery.
Conclusion

Grading/staging of a disease is often linked to therapeutic protocols in which quantitative imaging is playing an increasing role. Altered morphology can be measured as a change of size/shape or the change of pixel intensity on the radiographic image [28-30]. Since IVDD is thought to be a major factor that results lumbar pain, the follow-up of degenerative process is a common morphological aspect that should be considered. Consequently, several grading systems have been worked out in the past decades including those which apply MR images, primarily from T2W measurements. During the past couple of decades there have been several suggestions to standardize the diagnostic imaging scheme of IVDD. Thompson et al. made one of the first grading of IVDD based on gross anatomy [31]. This system was based on 15 human cadaveric spines and assumes that the process of disc degeneration advances at a similar pace in all discs, which is not always the case [32]. MR imaging classifications such as Schneiderman’s [1] and the modified Pfirrmann scheme designed for both elderly [6], and young subjects [33] provide also discontinuous scales while other classifications have been developed for more specific changes [34-37]. Today the most accepted grading system of IVDD is that of Pfirrmann et. al. which is a semi-quantitative grading based on the T2W signal changes of the nucleus pulposus and height of lumbar discs sorting them in five categories [4]. Despite of the justified morphological standards, the measurements of IVDD on T2-weighted images may hide important features because of numerous factors in signal detection [20, 38]. Consequently, the visual perception of signal intensity on T2-weighted images is problematic which may lead differences between raters, especially when the comparing observers have different level of experience and working area or the examiners show overall variation in stringency. These prior sources of disagreement will be reflected in dissimilarity between the observers [39]. Nevertheless, in most studies the inter-observer differences are usually settled by mutual consensus [38, 40, 41], or inviting a third observer [42, 43]. It is noteworthy to mention that the qualitative T2-weighted image of the lumbar spine shows only cross-sectional information of the T2 relaxation curve, while T2 mapping quantitatively describes the whole T2 relaxation process. Additionally, T2 relaxation time evaluation provide the most reliable correlation with changes of molecular environment in the disc [17], which consist of water, proteins, proteoglycan and collagen. Despite these facts, only a few studies have suggested that T2 relaxation time evaluation may have a role to determine the range of T2 values for disc degeneration [20]. In our study the possibility of quantitative classification in intervertebral disc degeneration (IVDD) was investigated: T2 cut-off values with regard to morphological classifications of Pfirrmann, Schneiderman and Pfirrmann type scheme were determined.

Similarly to others[16-18] our results showed that T2 relaxation time of lumbar intervertebral discs decreased with increasing morphological grades and age, likely reflecting a decrease in proteoglycan and water content [44, 45]. There was no significant difference in T2 values between Pfirrmann grades IV and V. This finding was considered
suggestive of decreased the disc high in grade V, which makes difficult to distinguish between grades IV and V. It is not the case in Schneiderman grading system, where the classification reflects alterations only in the signal intensity of the disc [1].

Only fair overall agreement was found between observers in both Pfirrmann and Schneiderman classification systems regarding the inter-observer reliability. In Pfirrmann and Pfirrmann type classification slight to moderate agreement was observed, while Schneiderman grading system showed fair to substantial agreement among grades. In agreement with other studies, disagreement was more frequent in Pfirrmann's grade III and IV [4], while Schneiderman's grades I and II showed the lowest kappa values in terms of inter-observer reliability, most likely due to the similar appearance of morphological alterations in the nucleus pulposus for both classifications.

Yet, the judgment of signal intensity on the T2-weighted images is subjective; as a consequence inter-observer differences exist. In fact, screening the literature it seems that observer variability of degenerative lumbar discs show a high variation depending on the backgrounds, the levels of experience and the investigated variable (Table 4) [4, 6, 18, 38, 46-52]. In this regard, the possibility to discriminating grades on morphological images is not always possible with full agreement.

Correspondingly, the use of receiver operating characteristic analysis between each grade of the IVDD yields quantitative T2 cut-off values. The cut-off points can be determined with an approximate reliability from the area under the curve (AUC) [20]. The AUC values in the present study were within moderate to high accuracy level (0.7-1). These data indicate that this quantitative grading is a more powerful way to evaluate IVDD, even if the inter-observer reliability is that low.

**Summary**

In conclusion, this study reflected that T2 values tended to decreased with increasing age or classification grade in nucleus pulposus mostly due to the decrease in water and proteoglycan content. Inter-observer agreement of morphological evaluation in patients with intervertebral disc degeneration was only fair on the classification of moderate degeneration level both in Pfirrmann and Schneiderman schemes. Therefore we calculated the boundaries of quantitative T2 classification based on morphological evaluation schemes. Based on our results, T2 cut-off values seems to be a more reliable method to define the degree of disc degeneration quantitatively, which may help staging the IVDD more accurately, even though the definitions of intervertebral disc degeneration (IVDD) in MRI are still not uniform.


**Personal Information**