CrossMark

ORIGINAL ARTICLE

Intervertebral disc height loss demonstrates the threshold of major pathological changes during degeneration

Joshua P. Jarman · Volkan Emre Arpinar ·
Dhiraj Baruah · Andrew P. Klein ·
Dennis J. Maiman · L. Tugan Muftuler

Abstract

Purpose Quantitative MRI techniques were utilized to study intervertebral disc degeneration. Main focus was to develop a novel approach to quantify disc height loss associated with disc degeneration. Currently there is no universally accepted metric of degeneration based on measurement of disc height. Such quantitative imaging methods would complement qualitative visual assessment methods currently used and offer a valuable diagnostic tool.

Methods 51 adult participants took part in this MRI study. T2 weighted images were used to obtain disc height index (DHI) and also a semi-quantitative metric based on relative voxel intensities. For DHI, each disc was given a score based on standard deviations from the mean DHI of healthy discs. Diffusion Weighted MRI was used to assess morphological changes in the nucleus pulposus. Conventional Pfirrmann classification was used as the gold standard to assess these quantitative approaches.

Results At deviations of up to 1.5σ below normative disc height, levels of apparent diffusion coefficient (ADC) and normalized T2 intensity were maintained. Once disc compression reached 1.5σ, there was a massive drop in ADC and normalized T2 intensity. Pfirrmann degeneration scores also increased after the 1.5σ mark.

Conclusions This study provides new, unbiased quantitative imaging tools to assess disc degeneration. We observed that these quantitative MRI measures indicate a threshold beyond which major pathological changes take place concurrently. Combined information from DHI, ADC and T2 images construct a set of novel biomarkers that could be used to identify degenerating discs that are approaching the threshold and possibly intervene before major pathologic changes occur.

Keywords Intervertebral disc degeneration · Apparent diffusion coefficient · Intervertebral disc height · Pfirrmann classification · Low back pain

J. P. Jarman
Medical Student, Medical College of Wisconsin, Milwaukee, WI 53226, USA

V. E. Arpinar · D. J. Maiman · L. Tugan Muftuler (✉)
Department of Neurosurgery, Medical College of Wisconsin, Froedtert Hospital, 9200 W. Wisconsin Ave., Milwaukee, WI 53226, USA
e-mail: lmuftuler@mcw.edu

D. Baruah · A. P. Klein
Department of Radiology, Medical College of Wisconsin, Milwaukee, WI 53226, USA

L. Tugan Muftuler
Center for Imaging Research, Medical College of Wisconsin, Milwaukee, WI 53226, USA

Introduction

Intervertebral disc (IVD) degeneration is a universal and natural process. As we age, our discs lose elasticity and begin to show signs of degeneration. Proteoglycans, a water retaining molecule, diminish within the degenerating discs. With the decrease in water, the discs lose their ability to resist compression and torque. Eventually the discs begin to lose height and structural integrity. The cascade of events that manifest as the discs degenerate can often progress to a pathologic state leading to chronic pain [1].

Physicians have routinely used magnetic resonance imaging (MRI) to assist in the clinical diagnoses of disc degeneration [2, 3]. Several MRI-based metrics have been

developed to facilitate communication between clinicians concerning the degree of degeneration in patients. The current gold standard is the 5-grade system developed by Pfirrmann et al. [3]. They demonstrated that the discs could be rated by independent raters with reasonably high confidence and reproducibility. While others have tried to expand and improve upon this technique [4], it is qualitative in nature and subject to bias. While most discs in Pfirrmann's study were rated in agreement, some of the ratings differed by one grade. In another rating, the difference was two grades. These errors could have a large impact on making the right decision for the course of treatment. Moreover, such generalized grading schemes might miss the fact that two discs rated with the same grade could actually be in different phases of pathologic changes. Therefore, it is imperative to discover biomarkers that supplement the Pfirrmann grading system.

Pfirrmann classification includes visual assessment of disc height and other features seen in T2 weighted (T2W) images. If it maintained a nominal height, it would be given a lower score, provided other aspects of the visual inspection complemented that finding. Several other studies have also documented an association between disc height loss and disc degeneration [3, 5–7]. However, there has not been any comprehensive analysis of the link between loss of structural integrity and various other physiological and morphological changes.

To address the need for such imaging biomarkers, we proposed a new method to quantify disc height loss and studied its association with degeneration. We also investigated the associations between disc height loss and changes in apparent diffusion coefficient (ADC) and voxel intensities in T2W images because there is strong evidence that disc height loss, restricted water diffusion in the nucleus pulposus (NP) and dehydration are good indicators of degeneration and those can be probed indirectly using MRI [5–12].

Materials and methods

Subjects and MRI protocol

51 adult participants took part in this study. The study was approved by the IRB and written consents were obtained from all participants. The demographic characteristics of the study cohort are given in Table 1.

Images of the lumbar spine were acquired using a 3T GE Discovery MR750 (Waukesha, WI, USA) MRI system. All images were acquired with a CTL-spine coil, FOV = 310 mm and 16-sagittal slices with 3 mm thickness. A conventional T2W MRI was acquired using a Fast Spin-Echo sequence with TR/TE = 4500 ms/104 ms,

Table 1 Demographic characteristic of the study cohort

	Age	Height (m)	Weight (kg)	BMI (kg/m^2)		N
Median	32	1.73	79	25.9	Female	19
Range	20–57	1.55–1.93	54–113	18.8–41.6	Male	32

ETL = 24 and 1 mm in-plane resolution. Diffusion weighted images (DWI) were acquired using a single-shot EPI with TR/TE = 2100 ms/75 ms, NEX = 12 and 2.4 mm in-plane resolution $b = 0, 600$ s/mm^2.

T2W images were reviewed by two expert radiologists and a trained medical student for Pfirrmann classification of the lumbar IVDs. If there was a disagreement between the raters, the median score was assigned to the disc.

Multi-parametric quantitative MRI of disc degeneration

ADC calculation

ADC is a measure of bulk diffusion of water molecules and there is compelling evidence in published data indicating its use for assessment of disc degeneration [8–11, 13]. It is an indirect measure of some pathologic changes in the NP such as fibrosis, cross-linking, and dehydration.

The ADC values were calculated in each voxel from the motion-corrected DWI images using Eq. 1:

$$S_1 = S_0 \cdot \exp(-b_1 \cdot \mathrm{ADC}) \tag{1}$$

where S_1 and S_0 are the signal intensities measured with and without the diffusion weighting, respectively. Regions of interest (ROI) that encompass the whole NP were drawn on the S_0 image and voxel-averaged ADC values were calculated for each IVD.

Normalized disc signal intensity in T2W images

T2W images reflect the progression of degeneration very well because the voxel intensities show good correlation with water and proteoglycan content in the IVDs [13, 14]. It is also used as one of the features in Pfirrmann classification. Therefore, we derived a semi-quantitative measure from the T2W images. First, the mean voxel intensity in each NP was calculated. Then another ROI was drawn in a uniform region of the gray matter of the spinal cord and the mean intensity was calculated. Then, the mean intensity in each disc was divided by the mean intensity in the spinal cord gray matter to obtain a metric that can be compared across subjects. Based on earlier reports [15], we expected that the inter-subject variations in T2W signal in spinal cord gray matter would be minimal in subjects with no spinal cord anomalies.
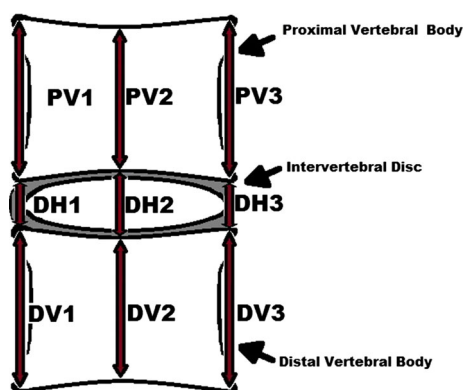
Fig. 1 A sketch of measurements taken from the vertebral bodies and IVDs for DHI calculation given in Eq. 2

Disc height index

Here we propose a new method to assess disc height loss with degeneration. Part of this method was modeled after a technique that was originally proposed by Lu et al. [16] and later refined by Masuda et al. [17]. They introduced a measurement called disc height index (DHI), which helps minimize the inter-subject variations in overall size of the vertebral column, enabling inter-subject analyses. The proximal (PV) and distal (DV) vertebral body height and IVD height (DH) were measured from the anterior (1), middle (2), and posterior (3) portions of each respective disc level on T2 W images (a sketch is shown in Fig. 1). The measurements were performed on the midsagittal slice. The corners of the vertebral bodies and the midpoints of the endplates were marked using an approach similar to the one described by Inoue et al. [18] and Quint et al. [19]. Then, the measurement lines were drawn between those landmarks as depicted in Fig. 1 and distance measurements were taken. DHI was determined using Eq. 2:

$$DHI = 2 * \frac{DH1 + DH2 + DH3}{(PV1 + PV2 + PV3) + (DV1 + DV2 + DV3)}. \quad (2)$$

Although DHI calculation given in Eq. 2 or similar methods proposed by Haefeli et al. [7] or Inoue et al. [18] are useful, it does not tell us how much a degenerating disc collapsed with respect to a healthy one. Therefore, we first established normative data for each lumbar IVD by picking discs with no discernible degeneration (grade 1 or 2) from the subjects. The DHI scores of all healthy discs at a particular level (e.g., L1/L2) were averaged so that each disc was assigned a mean and standard deviation, representing the healthy population (Table 2). Then, each disc of every subject was given a DHI_diff score based on how much its DHI deviated (in units of standard deviations, σ) from the

Table 2 Number (N), mean and standard deviation (σ) of DHI of healthy discs for each lumbar level

	L1/L2	L2/L3	L3/L4	L4/L5	L5/S1
N	42	40	34	31	28
Mean DHI	0.292	0.338	0.367	0.420	0.389
σ DHI	0.039	0.044	0.045	0.056	0.061

mean DHI of the corresponding healthy disc. The discs selected for this normative data included those from both sexes, different races and participants with different heights (155–193 cm, mean = 173 cm) and BMI (18.7–41.5, mean = 26.3). Therefore, the standard deviation represents the most probable variations due to these main factors. Further changes from this standard deviation might represent a pathologic condition.

Statistical analysis

SPSS software version 21 was used for statistical analysis (Armonk, NY, USA). Significance level of 0.05 was set for all analyses. First, one-way ANOVA was conducted to compare the difference between DHI scores with respect to Pfirrmann grades. Then, pairwise comparisons were done using independent-samples t test, assuming unequal variances.

Then, the associations between DHI scores and ADC values or normalized T2 intensity values were analyzed. Seven discs that did not have ADC values were excluded from this analysis (one subject did not undergo DWI scan and NP of two discs were not visible in S_0 images due to severe degeneration). To study the associations between DHI loss, Pfirrmann grades and ADC or T2 changes, the discs were binned based on their DHI scores (with 0.5σ increments) and average ADC, normalized T2W intensity and Pfirrmann grades were calculated for each bin. Then, the differences between group means were compared using one-way ANOVA test for ADC and T2 and Kruskal–Wallis test for Pfirrmann grades.

Results

DHI and Pfirrmann

Pfirrmann classification suggests that grade 2 and 3 discs should not have any discernible disc height loss, while grade 4 and 5 should show significant collapse. This visual assessment was first validated using a box plot of DHI_diff against Pfirrmann grades (Fig. 2). Height loss was not significant in grade 2 discs, but noticeable loss could be seen with grade 3 discs. Those discs had a median DHI_diff
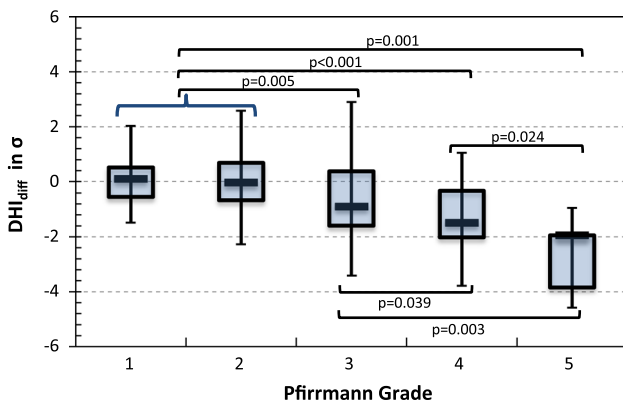
Fig. 2 *Box plot* of DHI$_{\text{diff}}$ (in units of standard deviation) plotted against Pfirrmann grades. Each *box plot* shows the median, 1st and 3rd quartiles and minimum and maximum values of DHI$_{\text{diff}}$ for each grade. One-way ANOVA analysis demonstrated significant differences in DHI$_{\text{diff}}$ with increasing Pfirrmann grades ($p < 0.001$)
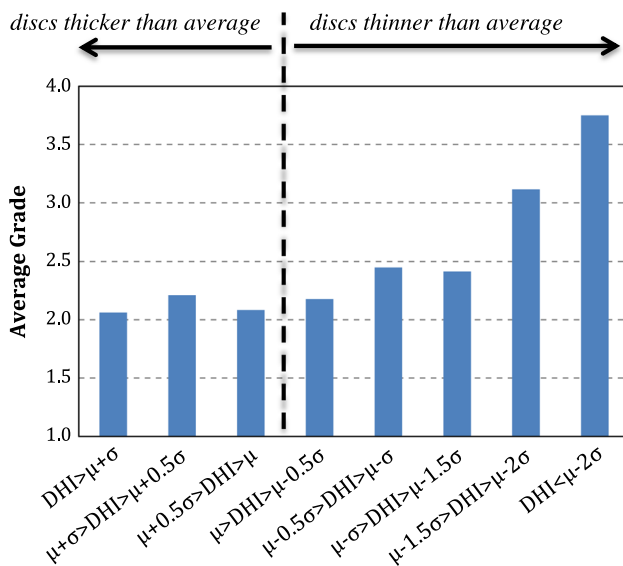

Fig. 3 Mean Pfirrmann grades for discs with different levels of DHI loss. For instance, DHI $<\mu-\sigma$ means that the discs in that bin is at least one standard deviation thinner than the mean (μ) of healthy discs. A significant difference in Pfirrmann grades was found between the discs with DHI $>\mu-1.5\sigma$ and DHI $\leq\mu-1.5\sigma$ ($p < 0.001$)

Fig. 4 Mean normalized T2W voxel intensities (in arbitrary units) for discs with different levels of DHI loss (in units of σ). The T2W signal drop after 1.5σ DHI loss was highly significant ($p < 0.001$)

Pfirrmann grades in each DHI bin. The resulting bar graph shows that the average Pfirrmann grades were relatively low (around 2.2) until DHI loss was more than 1.5σ (Fig. 3). At this threshold, a noticeable increase in average grade was observed. Pairwise comparisons of the first 6 bins (DHI $>\mu-1.5\sigma$) were done using Kruskal–Wallis test and no significant differences were found ($p = 0.223$). Then, discs with DHI $>\mu-1.5\sigma$ and DHI $\leq\mu-1.5\sigma$ were separated into two groups and compared using the same test and a significant difference was found ($p < 0.001$).

DHI and T2 intensity

Normalized T2W voxel intensity remained consistent at just under 0.23 (arbitrary units) until the DHI$_{\text{diff}}$ was about 1.5σ (one-way ANOVA, $p = 0.959$). After this cutoff point, however, there is a substantial drop in intensity (Fig. 4). A t test comparing discs above and below this cutoff demonstrated that the T2W signal drop after this threshold was highly significant ($p < 0.001$, without assuming equal variances).

DHI and ADC

A trend similar to T2W intensity changes was also observed for ADC values (Fig. 5). The ADC values in discs remained around 1.5×10^{-3} mm^2/s for DHI $>\mu-1.5\sigma$ (one-way ANOVA, $p = 0.976$). However, after the 1.5σ mark, a significant drop in ADC values was noted ($p = 0.001$).

of about one σ. Further decline can be seen with grade 4 and 5 discs. One-way ANOVA analysis demonstrated significant differences in DHI$_{\text{diff}}$ with increasing Pfirrmann grades ($p < 0.001$). Results of pairwise comparisons are shown in Fig. 2. Here, data from grade 1 and 2 discs were combined and compared to discs with higher grades since they are considered as healthy discs and their DHI values were not significantly different in our analysis ($p = 0.564$). Significant loss of disc height with increasing degree of degeneration was evident after grade 2.

The relation between Pfirrmann grades and disc height loss was further investigated by calculating the average
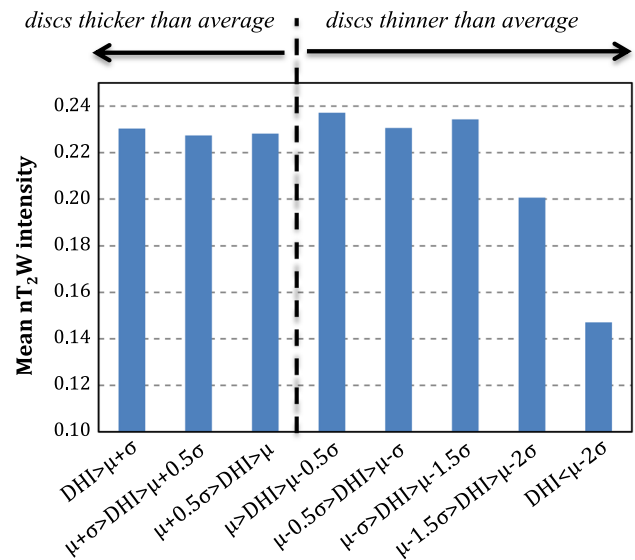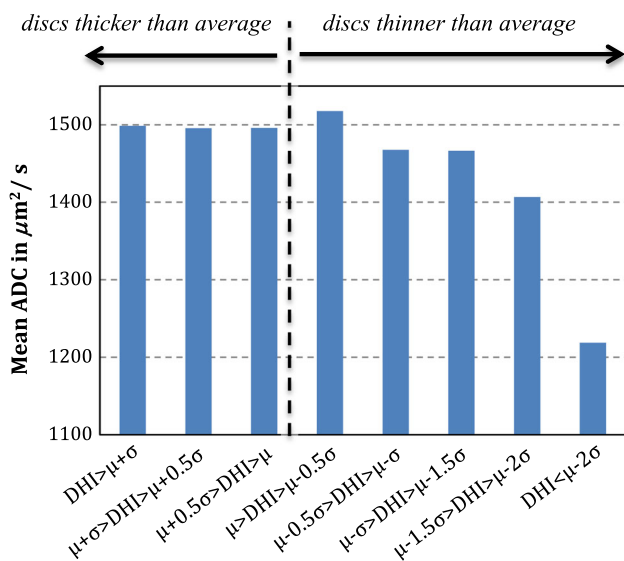
Fig. 5 Mean ADC in disc NP for discs with different levels of DHI loss. Similar to the trends observed for T2W voxel intensities, a significant drop in ADC values was noted after 1.5σ DHI loss ($p = 0.001$)

A summary of descriptive statistics for the quantitative MRI measures is given in Table 3.

Multi-parametric assessment of disc degeneration

To illustrate the potential of these MRI biomarkers in the clinical evaluation of disc degeneration, data from a few subjects were investigated individually. In those subjects, some discs were rated by the radiologists as normal or mildly degenerated (Pfirrmann grade 2 or 3), but the DHI_{diff}, ADC and normalized T2W voxel intensity values suggested that these discs already began to undergo significant pathologic changes. Data from one of those subjects is presented in Fig. 6 and Table 4. For instance, the L3/L4 disc was rated grade 3 but it had a lower ADC value, almost the same disc height loss and slightly higher T2W signal compared to the L4/L5 disc, which was graded as 4. Therefore, the imaging metrics indicated that the L3/L4 disc might be closer to a grade 4 disc, rather than a grade 3 disc (e.g., L2/L3 disc in the same subject).

Discussion

DHI and DHI_{diff} are novel biomarkers that could be instrumental in the evaluation of IVD degeneration. Accurate diagnoses depend on metrics that are easily reproducible, clear to interpret, and fast to perform. We propose that DHI fulfills these criteria. These new metrics assess height loss in vivo and correlates well with several pathological processes probed by other MRI techniques.

We observed a threshold of about 1.5σ, beyond which indicators of degenerative processes had accelerated. Therefore, the DHI, in combination with other MRI biomarkers, has the potential to aid the clinician assess the condition of degenerating discs more accurately.

Although Pfirrmann grading has greatly improved and standardized classification of degenerating discs, it is a subjective assessment and leads to unknown amount of bias. If Pfirrmann classification were supplemented by unbiased quantitative metrics, the physicians would be equipped with more accurate data to diagnose the condition of the discs and plan the best treatment regimen.

Of note, our study discovered discrepancies between Pfirrmann grades and the level of degeneration indicated by ADC, T2W voxel intensity, and DHI. In clinical practice, it would not be beneficial for a patient if his treatments were planned based on low Pfirrmann rating, although his discs were actually in further stages of degeneration. DHI is a potentially useful biomarker and it has a cutoff point beyond which discs undergo major degenerative changes.

Addition of other quantitative metrics might further improve the assessment of degenerating discs. Strong evidence supports the use of ADC and T2W voxel intensity in the assessment of disc health [8–11, 13, 14, 20]. Each metric reveals additional information about the pathological processes occurring inside the discs. Therefore, these measures add meaningful insight into the phase of ongoing degenerative processes.

The precedent for diagnostic data using standard deviations and a cut off for disease processes is already established with dual-energy X-ray absorptiometry (DXA) and osteoporosis. The mean bone density for a healthy 30-year old is determined and a t score in units of standard deviations demonstrates whether the bones are more or less dense than those of a 30-year old. At a t score of -1.0σ to -2.5σ signifies osteopenia and a score greater than -2.5σ signifies osteoporosis [14, 21]. In a similar fashion, a DHI score of 1.5σ might be used as a threshold of significant aberrations in degenerating discs. If feasibility of this approach is demonstrated with clinical outcomes in future studies, physicians would be equipped with a clear cutoff of disease process instead of ambiguous scores and grades.

In this study, care was taken to minimize potential sources of error and bias. Since Pfirrmann classification was used as the gold standard, discs were examined by three experienced raters and the median score of the raters were taken. ROI drawings could be another source of bias and error; nevertheless the technician was carefully trained to encompass anatomy of interest and drawings were verified by an independent operator. We preferred to use the six vertebral measurements approach proposed by Masuda

Table 3 Mean ± standard deviation for Pfirrmann grades, normalized T2 intensity and ADC (in μm^2/sec) within each DHI bin

	DHI > $\mu + \sigma$ (N = 32)	$\mu + \sigma$ > DHI > $\mu + 0.5\sigma$ (N = 38)	$\mu + 0.5\sigma$ > DHI > μ (N = 36)	μ > DHI > $\mu - 0.5\sigma$ (N = 45)	$\mu - 0.5\sigma$ > DHI > $\mu - \sigma$ (N = 29)	$\mu - \sigma$ > DHI > $\mu - 1.5\sigma$ (N = 29)	$\mu - 1.5\sigma$ > DHI > $\mu - 2\sigma$ (N = 26)	DHI < $\mu - 2\sigma$ (N = 20)
Grade	2.06 ± 0.56	2.21 ± 0.62	2.08 ± 0.77	2.18 ± 0.65	2.45 ± 1.02	2.41 ± 0.78	3.12 ± 1.07	3.75 ± 0.91
T2	0.230 ± 0.054	0.227 ± 0.057	0.228 ± 0.049	0.237 ± 0.059	0.231 ± 0.072	0.234 ± 0.061	0.201 ± 0.091	0.147 ± 0.049
ADC	1,499 ± 227	1,496 ± 247	1,496 ± 219	1,518 ± 234[a]	1,468 ± 304[b]	1,466 ± 292[a]	1,407 ± 291[b]	1,219 ± 264[a]

The number of discs in each bin is listed in parentheses

[a] 1 Entry missing

[b] 2 Entries missing

Fig. 6 T2W image of lumbar spine. Pfirrmann grades are labeled next to each disc

Table 4 MRI measurements from intervertebral discs shown in Fig. 5

Grade	σ below mean DHI	ADC (in μm^2/sec)	T2W
2	0.22	1,526	0.331
3	1.60	1,290	0.237
3	2.58	872	0.196
4	2.51	1,062	0.141
2	0.43	1,598	0.286

et al. [17] instead of the single vertebral measurement presented by Inoue et al. [18] because the average of six vertebral measurements might reduce small errors made in the placement of measurement lines (Fig. 1). A limitation of the study presented here is the difficulty in obtaining reliable measurements in severely degenerated discs. DHI loss could probably be estimated; however, T2W voxel intensity and ADC measurements could be compromised because it becomes harder to distinguish the compartments of the disc.

IVD degeneration is a process that affects all people [1]. Understanding the degeneration processes and specific conditions that might lead to low back pain would benefit millions of people suffering from this ailment. Combined information from imaging biomarkers could be used to identify the characteristics of disc degeneration better and plan effective treatments.

References

1. Roh JS, Teng AL, Yoo JU, Davis J, Furey C, Bohlman HH (2005) Degenerative disorders of the lumbar and cervical spine. Orthop Clin North Am 36:255–262
2. Miller TT (2004) Imaging of disk disease and degenerative spondylosis of the lumbar spine. Semin Ultrasound CT MR 25:506–522
3. Pfirrmann CW, Metzdorf A, Zanetti M, Hodler J, Boos N (2001) Magnetic resonance classification of lumbar intervertebral disc degeneration. Spine (Phila Pa 1976) 26:1873–1878
4. Griffith JF, Wang YX, Antonio GE, Choi KC, Yu A, Ahuja AT, Leung PC (2007) Modified Pfirrmann grading system for lumbar intervertebral disc degeneration. Spine 32:E708–E712
5. Benneker LM, Heini PF, Anderson SE, Alini M, Ito K (2005) Correlation of radiographic and MRI parameters to morphological and biochemical assessment of intervertebral disc degeneration. Eur Spine J 14:27–35
6. Boos N, Weissbach S, Rohrbach H, Weiler C, Spratt KF, Nerlich AG (2002) Classification of age-related changes in lumbar intervertebral discs: 2002 Volvo award in basic science. Spine (Phila Pa 1976) 27:2631–2644
7. Haefeli M, Kalberer F, Saegesser D, Nerlich AG, Boos N, Paesold G (2006) The course of macroscopic degeneration in the human lumbar intervertebral disc. Spine 31:1522–1531
8. Antoniou J, Demers CN, Beaudoin G, Goswami T, Mwale F, Aebi M, Alini M (2004) Apparent diffusion coefficient of intervertebral discs related to matrix composition and integrity. Magn Reson Imaging 22:963–972
9. Kealey SM, Aho T, Delong D, Barboriak DP, Provenzale JM, Eastwood JD (2005) Assessment of apparent diffusion coefficient in normal and degenerated intervertebral lumbar disks: initial experience. Radiology 235:569–574
10. Kerttula LI, Jauhiainen JP, Tervonen O, Suramo IJ, Koivula A, Oikarinen JT (2000) Apparent diffusion coefficient in thoracolumbar intervertebral discs of healthy young volunteers. J Magn Reson Imaging 12:255–260
11. Kurunlahti M, Kerttula L, Jauhiainen J, Karppinen J, Tervonen O (2001) Correlation of diffusion in lumbar intervertebral disks with occlusion of lumbar arteries: a study in adult volunteers. Radiology 221:779–786
12. Ohmori K (2010) Clinical application of a novel assessment for lumbosacral stability. J Neurosurg Spine 13:276–282
13. Kerttula L, Kurunlahti M, Jauhiainen J, Koivula A, Oikarinen J, Tervonen O (2001) Apparent diffusion coefficients and T2 relaxation time measurements to evaluate disc degeneration. A quantitative MR study of young patients with previous vertebral fracture. Acta Radiol 42:585–591
14. Kanis JA, Delmas P, Burckhardt P, Cooper C, Torgerson D (1997) Guidelines for diagnosis and management of osteoporosis. The European foundation for osteoporosis and bone disease. Osteoporos Int 7:390–406
15. Smith SA, Edden RA, Farrell JA, Barker PB, Van Zijl PC (2008) Measurement of T1 and T2 in the cervical spinal cord at 3 tesla. Magn Reson Med 60:213–219
16. Lu DS, Shono Y, Oda I, Abumi K, Kaneda K (1997) Effects of chondroitinase ABC and chymopapain on spinal motion segment biomechanics. An in vivo biomechanical, radiologic, and histologic canine study. Spine (Phila Pa 1976) 22:1828–1834 (discussion 1834)
17. Masuda K, Aota Y, Muehleman C, Imai Y, Okuma M, Thonar EJ, Andersson GB, An HS (2005) A novel rabbit model of mild, reproducible disc degeneration by an anulus needle puncture: correlation between the degree of disc injury and radiological and histological appearances of disc degeneration. Spine (Phila Pa 1976) 30:5–14
18. Inoue H, Ohmori K, Miyasaka K, Hosoe H (1999) Radiographic evaluation of the lumbosacral disc height. Skeletal Radiol 28:638–643
19. Quint DJ, Tuite GF, Stern JD, Doran SE, Papadopoulos SM, Mcgillicuddy JE, Lundquist CA (1997) Computer-assisted measurement of lumbar spine radiographs. Acad Radiol 4:742–752
20. Mcauliffe MJ, Lalonde FM, Mcgarry D, Gandler W, Csaky K, Trus BL (2001) Medical image processing, analysis and visualization in clinical research. Proceedings. In: 14th IEEE Symposium on computer-based medical systems. pp 381–386
21. Kanis JA, Melton LJ, Christiansen C, Johnston CC, Khaltaev N (1994) The diagnosis of osteoporosis. J Bone Miner Res 9:1137–1141