Variability of T2-Relaxation Times of Healthy Lumbar Intervertebral Discs is More Homogeneous within an Individual Than across Healthy Individuals

A. Sharma, R.E. Walk, S.Y. Tang, R. Eldaya, P.J. Owen and D.L. Belavy

AJNR Am J Neuroradiol 2020, 41 (11) 2160-2165
doi: https://doi.org/10.3174/ajnr.A6791
http://www.ajnr.org/content/41/11/2160
Variability of T2-Relaxation Times of Healthy Lumbar Intervertebral Discs is More Homogeneous within an Individual Than across Healthy Individuals

A. Sharma, R.E. Walk, S.Y. Tang, R. Eldaya, P.J. Owen, and D.L. Belavy

ABSTRACT

BACKGROUND AND PURPOSE: When one uses T2 relaxometry to classify lumbar intervertebral discs as degenerated, it is unclear whether the normative data should be based on other intervertebral discs from the same individual or from a pool of extraneous controls. This study aimed to explore the extent of intra- versus intersubject variation in the T2 times of healthy intervertebral discs.

MATERIALS AND METHODS: Using prospectively acquired T2-relaxometry data from 606 intervertebral discs in 101 volunteers without back pain (47 men, 54 women) in a narrow age range (25–35 years), we calculated intra- and intersubject variation in T2 times of intervertebral discs graded by 2 neuroradiologists on the Pfirrmann scale. Intrasubject variation of intervertebral discs was assessed relative to other healthy intervertebral discs (Pfirrmann grade, ≤2) in the same individual. Multiple intersubject variability measures were calculated using healthy extraneous references ranging from a single randomly selected intervertebral disc to all healthy extraneous intervertebral discs, without and with segmental stratification. These variability measures were compared for healthy and degenerated (Pfirrmann grade ≥3) intervertebral discs.

RESULTS: The mean T2 values of healthy (493/606, 81.3%) and degenerated intervertebral discs were 121.1 ± 22.5 ms and 91.5 ± 18.6 ms, respectively (P < .001). The mean intrasubject variability for healthy intervertebral discs was 9.8 ± 10.7 ms, lower than all intersubject variability measures (P < .001), and provided the most pronounced separation for healthy and degenerated intervertebral discs. Among intersubject variability measures, using all segment-matched healthy discs as references provided the lowest variability (P < .001).

CONCLUSIONS: Normative measures based on the T2 times of healthy intervertebral discs from the same individual are likely to provide the most discriminating means of identifying degenerated intervertebral discs on the basis of T2 relaxometry.

ABBREVIATIONS: I, X, Xs, Xc, Xcs = various measures of variability in T2 times of intervertebral discs; IVD = intervertebral disc; np = used as a suffix to indicate measures assumed to represent nucleus pulposus.

L oss of signal intensity of the central part of intervertebral discs (IVDs) on T2-weighted MR images is a well-established indicator of underlying degenerative changes.1–4 Visual recognition of the extent of this loss of signal intensity also forms the basis of grading the extent of disc degeneration on frequently used grading schemes such as one proposed by Pfirrmann et al.5 To overcome the subjectivity of visual assessment, T2 relaxometry provides a reliable, objective, and continuous measure of IVD health, the T2 time.6–9 Previous studies have used this technique to identify differences in IVDs from cohorts differing in their Pfirrmann grade, age, and presence of back pain.7,9–12

Theoretically, T2 relaxometry should be able to indicate pathologic changes in the IVDs before these become evident to human readers. However, this inherently requires a definition of a normative range against which the T2 time of a given IVD should be measured. At present, categorization of a given IVD into healthy or degenerated cannot be done with reasonable certainty simply on the basis of its T2 time. The existing literature not only lacks a normative definition, but it also fails to establish whether such a definition should be based on other IVDs from the same individual or from a pool of extraneous controls. While it has been shown that the segmental level influences the T2 times of the IVDs,13 it remains unclear whether these intersegmental variations between cohorts of IVDs from different subjects are larger or smaller than...
the intrasubject variation in T2 times between lumbar IVDs at different segmental levels.

In this study of pain-free volunteers within a narrow age range (25–35 years) who were scanned on the same scanner using identical scan parameters, we aimed to explore the extent of intrasubject-versus-intersubject variation in T2 times of healthy IVDs.

**MATERIALS AND METHODS**

**Participants**

This study represents a secondary analysis of the data collected for a previously reported study performed on 101 participants (47 men, 54 women; age range, 25–35 years) without spine disease. Exclusion criteria included current spinal pain, history of spinal surgery, history of traumatic injury to the spine, known scoliosis for which prior medical consultation was sought, current or prior smoker, known claustrophobia, and possible pregnancy.14–16

**MR Imaging**

All studies were performed on the same 3T scanner (Ingenia; Philips Healthcare). For T2 mapping, a sagittal spin-echo multiecho technique was used with 8 TEs (15.75, 36.75, 57.75, 78.75, 99.75, 120.75, 141.75, and 162.75 ms). Additional scanning parameters were a section thickness of 3 mm, an interslice distance of 1.5 mm, TR of 2600 ms, and TE of 70 ms.

**Image Analysis**

Two board-certified neuroradiologists with 1 and 16 years of practice experience independently evaluated sagittal T2-weighted images and graded T12 through S1 IVDs in each participant according to the Pfirrmann grading system.5 Differences were resolved by consensus. Only healthy (Pfirrmann grade 2 or 3) or degenerated (Pfirrmann grade ≥3) IVDs accepted by both neuroradiologists were included for subsequent analysis for this study.

For T2 mapping, each IVD was initially segmented manually on each image using the native polygon selections tool in the off-line software ImageJ (National Institutes of Health). A custom plug-in (ROI Analyzer; https://sites.google.com/site/daniellbelavy/home/roianalyser) was then used to calculate the T2 time in 5 anterior-posterior regions across the IVD in each sagittal image. These values were then interpolated to a fixed-sized IVD so that each IVD, when viewed in an axial plane, could be represented by a grid with 11 (left-to-right) columns (C1–C11) and 5 anterior-to-posterior rows (R1–5). To exclude the influence of the outer annulus fibrosus, we used the mean T2 time of the central 21 regions (shaded yellow) indicated by rows R2–4 and columns C3–9 to represent the T2 time of any given disc. Of these 21 regions, 11 regions with the highest T2 values were assumed to represent the core of the disc (asterisk), which would be least affected by the inner annulus fibrosus. Mean T2 value of these 11 pixels was assumed to represent the mean T2 time of the nucleus pulposus of the disc.

**FIG 1.** Artistic rendering of a fixed-sized intervertebral disc to which T2 values of all individual discs were interpolated. For each disc, T2 values of 55 equally sized regions were available that could be represented by a grid with 11 (left-to-right) columns (C1–C11) and 5 anterior-to-posterior rows (R1–5). To exclude the influence of the outer annulus fibrosus, we used the mean T2 time of the central 21 regions (shaded yellow) indicated by rows R2–4 and columns C3–9 to represent the T2 time of any given disc. Of these 21 regions, 11 regions with the highest T2 values were assumed to represent the core of the disc (asterisk), which would be least affected by the inner annulus fibrosus. Mean T2 value of these 11 pixels was assumed to represent the mean T2 time of the nucleus pulposus of the disc.
between the T2 (or T2np) time of a given disc and the average T2 (or T2np) times of all other IVDs with same Pfirrmann grade as that disc. Xs and Xsnp represented similar calculations restricted to the same segmental level as a given disc. Additional measures assessed intersubject variation with respect to a single randomly selected healthy IVD from the same segmental level or an average of 6 randomly selected healthy IVDs representing the T12–S1 segments (Table).

**Statistical Analysis**

Interobserver reliability of Pfirrmann grading was tested for the 2 readers, and the Cohen $\kappa$ was calculated. Mean and SDs were calculated for each measure of T2 variability. A paired 2-tailed Student $t$ test was used to assess differences between various measures of variability. A nonpaired Student $t$ test was used to assess differences between the variability measures of IVDs that were categorized as healthy and degenerated on the basis of the Pfirrmann grading system. A $P$ value $<.05$ was considered significant. A Bonferroni correction was applied to correct for multiple comparisons.

**RESULTS**

Of 606 IVDs, 493 (81.4%) were deemed healthy after a consensus read by 2 radiologists and were included for subsequent analysis. Of these, 489/493 (99.2%) were assigned grade 2 on the Pfirrmann scale. Individual reader agreement was excellent ($\kappa = 0.84$; 95% CI, 0.78–0.89) for categorization of IVDs as healthy or degenerated and good ($\kappa = 0.73$; 95% CI, 0.67–0.79) for a specific Pfirrmann grade. Readers disagreed on the grading of 61/606 (10.1%) IVDs. Readers were able to reach a consensus on all except 2 IVDs, which were excluded from subsequent analysis.

Mean T2 and T2np values for all IVDs were 115.5 ± 24.6 ms and 131.5 ± 30.0 ms, respectively. IVDs graded healthy (Pfirrmann grades 1 and 2) had higher T2 (121.1 ± 22.5 ms) and T2np (138.6 ± 26.8 ms) values compared with corresponding T2 and T2np values of 91.5 ± 18.6 ms and 101.0 ± 23.3 ms, respectively, for degenerated IVDs ($P < .001$ for both).

Intrasubject variability measures (I and Inp) were significantly lower than any of the corresponding intersubject variability measures ($P < .001$ for all, Table and Fig 2). Inp, while being significantly lower than all of the intersubject variability measures ($P < .001$ for all), was significantly higher than I ($P < .001$).

Intersubject variability was higher when a single or 6 randomly selected IVDs were used as a comparison rather than all healthy IVDs in all participants without or with stratification for the segmental level ($P < .001$ for all, Table and Fig 2).

Stratification based on the segmental level did not impact the variability, with no significant difference observed between X and Xs ($P = .12$) or between Xnp and Xsnp ($P = .27$).

**DISCUSSION**

Many previous studies have suggested that T2-relaxometry can provide a reliable, objective, and continuous quantitative measure of the health of lumbar IVDs. Despite these advantages, this technique has failed to replace traditional subjective assessment of the signal intensity of IVDs on T2-weighted images for categorization of a given IVD as healthy or degenerated. The current literature is also deficient in providing either a T2 value or a measure based on such a value that could be used to determine normalcy of a given IVD. Our study was not designed to develop such a measure. Instead, it aimed to take the first step toward establishing the most appropriate reference standard. By showing that the T2 time of a given healthy IVD most closely matched that of other healthy lumbar IVDs from the same individual (Fig 2), our results suggest that other healthy IVDs within an individual, if available, are likely to provide the most optimal basis of the definition of normal against which a given IVD should be compared. This suggestion is further highlighted because the differences between healthy and degenerated IVDs were most stark when compared with internal rather than extraneous healthy IVDs (Fig 2). Using internal healthy IVDs as a reference would also have the advantage of circumventing the potential of variation in measured T2 times of discs scanned on different scanners.

A recent study highlighted the role of level stratification in MR imaging–quantification studies using T2 data. Our analysis of the same data demonstrated that while the level of stratification might be important when cohorts of IVDs are being compared, T2 times of healthy IVDs at other levels in the same individual are likely to provide a better measure of the health of a given IVD.
than T2 times of a healthy IVD at the same segmental level from any other given individual or even an average T2 time of a large number of the same segmental-level IVDs from many other healthy controls. The reason may be both due to genetic similarities between IVDs from the same individual and the fact that placement of any given individual within the MR imaging scanner is likely to produce somewhat unique subtle variations in the magnetic environment that may affect T2 calculations even under identical scanning equipment and parameters. In addition, comparison with other IVDs within the same scan can help overcome many other confounding factors such as disc hydration status, time of the day, and loading of the spine, which would be difficult to control in comparison across individuals but are known to affect the T2 time of IVDs.17,18

Of all comparison scenarios studied, T2 values of any given healthy IVD differed most from a randomly selected healthy IVD from a different individual but from the same segmental level. It is likely that the variation in T2 times in healthy IVDs in the population are a combination of both the underlying true differences in T2 times of healthy IVDs across subjects and systematic effects due to the noise encountered during the process of measuring the T2 times using MR imaging. While the true underlying variability in T2 times cannot be changed, a greater number of samples can minimize the contributions of the noise. This feature likely explains why the interobserver variability decreased as the number of control discs increased from 1 (Xcs) to 6 (Xc), with a further significant reduction when a much larger number of control discs were used, as was the case for calculation of X and Xs (Table). This benefit of using a larger number of IVDs to define the T2 value of healthy IVDs remained irrespective of whether the IVDs were used from the same segmental level or not. However, despite a relatively small number of intrasubject control discs, such discs provide a T2 value that may be expected to be most similar to that of a given healthy IVD.

Certain methodologic details of our experiment merit explanation of the rationale. As opposed to measuring the T2 time of the entire IVD, as in a previous study,13 we chose to investigate that of a central aspect of the IVD, deliberately trying to exclude the outer annulus. While the central portion of the IVD is known to undergo loss of T2 signal in the presence of pathology, the T2 time of the abnormal annulus (annular fissures) may be expected to increase. Given that annular fissures often accompany and perhaps precede the appearance of IVD desiccation,19,20 we propose that given the opposing effects of degenerative changes in these components of IVDs, it would be reasonable to exclude the outer annulus from T2 calculations if the T2 value is to be used to define the degenerated status of the disc. For the sake of simplicity, some previous studies have used a central hyperintense zone of IVDs as being representative of the nucleus pulposus.9,10,21 The nucleus pulposus is not readily identifiable on T2-weighted MR imaging as a distinct structure from the inner annulus.22 While the central hyperintense region on T2-weighted images that represents a combination of the nucleus pulposus and the inner annulus fibrosus shows a loss of signal intensity with disc

---

**FIG 2.** Plots of differences of T2 time in milliseconds (y-axis) of 604 lumbar IVDs with Pfirrmann grades of ≤2 (blue), 3 (red), and 4 (green) relative to mean T2 time (or T2 time when single) of other healthy lumbar IVDs from same individual (I), all other healthy IVDs in 101 study participants without (X) and with (Xs) stratification for the segmental level, 6 randomly selected healthy IVDs representing T12–S1 segments (Xc), and a single randomly selected IVD from same segmental levels (Xcs). Horizontal bars show means with standard errors. The variation in T2 times of healthy IVDs, as represented by the spread along the y-axis, is least for healthy IVDs with Pfirrmann grade of ≤2. Notice that the measure I provides the most and the measure Xcs provides the least discrimination between healthy and degenerated IVDs with Pfirrmann grades of 3 or 4, as indicated by separation of means along the y-axis.
degeneration in its entirety, it is possible that there are underlying subtle differences in the rate of signal loss in these 2 components of IVDs that are not fully understood. Assuming that the nucleus pulposus might have inherently higher T2 times given its higher hydration level relative to that of inner annulus fibrosus, which slowly decreases from the central to outer aspect of the inner annulus, we explored the possibility that a smaller number of regions with higher intensity might be more representative of the nucleus pulposus of the IVDs (Fig 1). Accordingly, we analyzed variability only on the basis of regions skewed toward higher intensity. Notably, variability increased when a smaller number of regions from the IVDs were used (Table).

The mechanisms for this consistent difference remain unclear. It is possible that an increase in variation when dealing with a smaller sample of pixels is simply a function of the increasing influence of noise that would be expected to be minimized by averaging a higher number of pixels. It is also possible that this amplified variation results from the fact that the number of regions selected as possible representations of the nucleus pulposus was arbitrary and not necessarily restricted to the size of the nucleus pulposus, which itself is difficult to establish. Inclusion of the nucleus pulposus and variable parts of the inner annulus fibrosus in T2 calculations, therefore, could have resulted in a higher degree of variation observed in our study. When it becomes technologically feasible to allow a reliable segmentation of the nucleus pulposus, it would be interesting to study the variation among T2 values of the nucleus pulposus alone in healthy lumbar IVDs.

Our study has some limitations. First, it was restricted to participants in a narrow age range. While this was critical in allowing us to test our hypothesis free from the confounding effects of age, it does not ensure generalizability of these results to other age groups. It is known that T2 times of IVDs are affected by age. In the absence of any previous studies indicating segmental variations in these effects, we expect our results to be similar in other age groups. Additional studies would be needed however to test this expectation. Second, the area of IVDs taken into consideration to represent IVDs free of the outer annulus was somewhat arbitrary. However, by demonstrating similar results even when the analysis was restricted to central pixels skewed toward higher T2 values, we think that our results were able to overcome this limitation. While all the participants were free of back pain, we do not think this to be a significant limitation of the study. Previous studies have shown that despite a varying burden of overall disc degeneration in individuals with varying predispositions, disc degeneration remains a disc-specific process that follows a remarkably similar natural history irrespective of the presence of symptoms. Furthermore, despite the absence of back pain, a number of IVDs in our patient population demonstrated evidence of overt disc degeneration as indicated by their Pfirrmann grades.

CONCLUSIONS

By demonstrating a significantly higher variation in the T2 times of IVDs across subjects, our study suggests that normative measures based on the T2 times of healthy lumbar IVDs from the same individual are likely to provide the most discriminating means of identifying degenerated IVDs on the basis of T2 relaxometry. If such a measure could be developed on the basis of a relatively small number of healthy IVDs, T2 relaxometry has the potential to become valuable not only for comparisons of cohorts but also as a reliable and objective means of identifying early degeneration of individual IVDs. While using a large pool of extraneous discs would be the next-best option, such a measure is likely to lack widespread utility due to potential variations in T2 quantification on different scanner types. Further studies are needed to ensure that these results remain valid across different age groups.

REFERENCES