How Should We Grade Cervical Disk Degeneration? A Comparison of Two Popular Classification Systems

Lukas Urbanschitz[1], Susanne Bensler[2], Sascha Merat[2], Christopher G. Lenz[3] and Karim Eid[3]

1) Department of Spine Surgery, Cantonal Hospital Baden, Baden, Switzerland
2) Department of Radiology, Cantonal Hospital Baden, Baden, Switzerland
3) Department of Orthopaedics, Cantonal Hospital Baden, Baden, Switzerland

Abstract:

Introduction: Despite being originally developed for the evaluation of lumbar disk degeneration, the Pfirrmann classification has emerged as the most popular classification system for cervical disk degeneration. However, with the Suzuki classification, a new classification system that is specifically tailored for the evaluation of cervical disk disease was introduced. In this study, we aim to evaluate differences in inter- and intraobserver reliability of both classifications in a head-to-head comparison.

Methods: In total, we have evaluated 120 cervical disks within 40 patients via magnetic resonance imaging according to the Pfirrmann and Suzuki classification. The degree of disk degeneration was evaluated by two independent musculoskeletal radiologists. After 6 months, the classification was reassessed to evaluate the intraobserver reliability. The inter- and intraobserver reliabilities were then calculated using Cohen's kappa.

Results: The inter- and intraobserver reliability provided a significant agreement between all ratings in Pfirrmann as well as the Suzuki classification ($p>0.001$). The interobserver reliability was determined to be fair in both the Suzuki classification ($\kappa=0.290$) and the Pfirrmann classification ($\kappa=0.265$). The intraobserver reliability was substantial in the Suzuki classification ($\kappa=0.798$), while it was almost perfect in the Pfirrmann classification ($\kappa=0.858$).

Conclusions: Although not designed for the evaluation of cervical disk degeneration, the Pfirrmann classification yielded equal inter- and higher intraobserver reliability. Both classification systems are viable options for the grading of cervical disk degeneration. While the Pfirrmann classification has the advantage of being better established, the Suzuki classification may be clinically superior due to a better representation of cervical disk degeneration and the consideration of disk bulging for the classification of cervical disk degeneration.

Keywords:

Pfirrmann, Suzuki, cervical disk degeneration, cervical classification, reliability

Spine Surg Relat Res 2021; 5(6): 347-351
dx.doi.org/10.22603/ssrr.2021-0048

Introduction

Cervical disk degeneration is part of the natural history of the cervical spine[1]. With increasing age, cervical degeneration is also expected to progress[2]. Although degenerative changes can be observed in asymptomatic individuals, cervical disk degeneration can lead to radiculopathy and myelopathy[3] and ultimately to cervical spine surgery[4].

To quantify disk degeneration, several classification systems, including histological[5] to radiological[6], have been developed. Since the introduction of the Pfirrmann classification[7] in 2001, it has been considered the gold standard for the evaluation of spinal disk degeneration. It was originally developed for the grading of lumbar disk degeneration. Nevertheless, it has become the most commonly used classification system in the literature even for cervical spine degeneration[8-11]. Other authors were able to demonstrate nearly perfect reliability for the assessment of cervical disk degeneration[12]. In 2017, the Suzuki classification[13] was introduced as a classification system specifically designed for the evaluation of cervical disk degeneration. As the Suzuki classification is rather new, quotes in literature are currently still

Corresponding author: Lukas Urbanschitz, Lukas_urbanschitz@hotmail.com
Received: March 9, 2021, Accepted: March 24, 2021, Advance Publication: April 28, 2021

scarce although expected to be rising[14,15]. For clinicians, each classification system can yield different uses and advantages for certain study questions. As both classifications are designed for the same pathology, it remains unclear in which areas the Suzuki classification can supersede a popular grading system like the Pfirrmann classification. It was, therefore, the aim of this study to evaluate if one of the classifications provided superior inter- and intraobserver reliability in a head-to-head comparison.

Materials and Methods

Study population

In total, we have evaluated 120 cervical disk levels (C3/4, C4/5, and C5/6) within 40 patients. All patients were retrospectively enrolled in this study. All MRI scans were made between November 2019 and December 2019. Only patients older than 18 years of age were included in this study. Patients with spinal tumors, traumatic injuries of the cervical spine, and previous spinal surgeries were excluded. This research has been approved by the IRB of the authors' affiliated institution.

Grading of cervical disk degeneration

Cervical disk degeneration was evaluated according to the Pfirrmann[7] and Suzuki[13] classification by two musculoskeletal radiologists. One of the radiologists repeated all measurements 6 months after the initial assessment for the evaluation of the intraobserver reliability. The MRI protocol available for every patient included T2-weighted sagittal, T1-weighted sagittal, T2 short T1 inversion recovery (STIR) sagittal sequences. All images were obtained using a 1.5-Tesla (T) scanner (Siemens Aera and Siemens Avanto Fit, Erlangen, Germany).

The pfirrmann classification

According to the Pfirrmann[7] classification, cervical disk degeneration is graded as follows: For grade I, disk is defined as a homogenous bright white disk with a clear distinction between the nucleus and annulus, with the nucleus being isointense to cerebrospinal fluid. For grade II, the main difference is the presence of an inhomogeneous nucleus with or without horizontal bands. For grade III, the nucleus becomes inhomogeneous and gray, and the distinction between the annulus and nucleus becomes unclear. The signal intensity becomes intermediate, and there can be a decrease in terms of disk height. The grade IV disk degeneration is defined as a gray to black disk; thus, it would be hard to distinguish nucleus from annulus. The disk appears hypointense, and the height can be moderately decreased. Lastly, grade V is the highest grade of disk degeneration; it is characterized by a collapsed disk space.

The Suzuki classification

According to the Suzuki[13] classification, cervical disk degeneration is graded as follows: For grade 0, disk is found to be healthy, which is characterized as a disk with a homogenous and high intensity nucleus without height loss. Grade I is characterized as a nucleus that is inhomogeneous without disk bulge or height loss. The grade II disk is characterized as a disk without clear annulus and nucleus distinction, and disk bulge is also noted. However, decrease in height must be less than 25%. Lastly, grade III disk degeneration is defined as a disk with a height loss of over 25%.

Statistical analysis

Inter- and intraobserver reliability was evaluated using Cohen's kappa. Observer reliability was rated as proposed by Landis and Koch[16] as <0.00 for poor, 0.00-0.20 as slight, 0.21-0.40 as fair, 0.41-0.60 as moderate, 0.61-0.80 as substantial, and 0.81-1.00 as almost perfect. To compare two individual variables, Spearman rank-order correlation was employed. A p-value of less than 0.05 was considered significant. Statistical analysis was performed on SPSS Statistics version 23 (2015, IBM, Chicago, Illinois, United States).

Results

In total, 40 cervical MRI scans were included in this study. The cohort consisted of 24 female and 16 male patients, with a mean age of 58 years (SD 19.5; range 22-88 years). The distribution of cervical disk degeneration according to the rater is depicted in Table 1, 2.

There was also a significant intrarater reliability ($p<0.001$) found after the reevaluation 6 months after the initial assessment. The Pfirrmann classification yielded almost perfect intrarater reliability ($\kappa=0.858$), while the Suzuki classification provided substantial intrarater reliability ($\kappa=0.798$). When analyzing the ratings for the Pfirrmann classification of the first observer, a grade I Pfirrmann disk degeneration was not found in neither the first nor the second rating. One disk at the level C4/5 war rated as grade II in the first and as grade III in the second rating. Of the 56 disks that were rated as grade III, 5 disks were classified as grade IV in the second rating. Out of the 51 disks that were classified as grade IV in the first rating, 47 disks received the same grade, and 4 disks were classified as grade III in the second rating. There was a perfect agreement on all 12 disks that were rated as grade V according to Pfirrmann. In the ratings of the Suzuki classification, one disk was rated as grade 0 in the first rating but as grade I in the second rating. Of those 29 disks that were classified as Suzuki grade I in the first rating, 6 were classified as grade II instead. Of the 59 disks that were classified as grade II, 3 were classified as grade I in the second rating. For the disks that were classified as grade III according to Suzuki classification, four were rated as grade II in the second observation.

A significant ($p<0.001$) interrater reliability was noted between the radiologists. However, the interobserver agreement was fair in both Pfirrmann ($\kappa=0.265$) and Suzuki ($\kappa=0.29$). When analyzing the differences between the observers, the

Table 1. Pfirrmann Classification According to Each Cervical Level and Rater.

	Pfirrmann I	Pfirrmann II	Pfirrmann III	Pfirrmann IV	Pfirrmann V
Rater 1a C3/4	0	0	22	17	1
Rater 1b C3/4	0	0	21	18	1
Rater 2 C3/4	2	8	13	16	1
Rater 1a C4/5	0	1	18	18	3
Rater 1b C4/5	0	0	18	19	3
Rater 2 C4/5	1	9	12	15	3
Rater 1a C5/6	0	0	16	16	8
Rater 1b C5/6	0	0	17	15	8
Rater 2 C5/6	2	6	11	9	12
Rater 1a Total	0	1	56	51	12
Rater 1b Total	0	0	56	52	12
Rater 2 Total	5	23	36	40	16

The ratings are depicted as the first (1a) and second (1b) rating of the first observer and the rating of the second (2) observer.

Table 2. Suzuki Classification According to Each Cervical Level and Rater.

	Suzuki 0	Suzuki I	Suzuki II	Suzuki III
Rater 1a C3/4	0	12	19	9
Rater 1b C3/4	0	10	22	8
Rater 2 C3/4	2	14	13	11
Rater 1a C4/5	1	7	23	9
Rater 1b C4/5	0	9	23	8
Rater 2 C4/5	1	16	9	14
Rater 1a C5/6	0	10	17	13
Rater 1b C5/6	0	8	20	12
Rater 2 C5/6	2	14	6	18
Rater 1a Total	1	29	59	31
Rater 1b Total	0	27	65	28
Rater 2 Total	5	44	28	43

The ratings are depicted as the first (1a) and second (1b) rating of the first observer and the rating of the second (2) observer.

single disk that was rated as grade II according to the Pfirrmann classification was graded as grade III by the second observer. Of the 56 disks that were rated as grade III according to Pfirrmann, the second observer agreed only in 20 disks and classified the rest as follows: 5 as grade I, 19 as grade II, and 12 as grade IV. Of the 51 disks classified as grade IV, the second observer agreed in most cases (27), but classified the rest as follows: 2 disks as grade II, 15 disks as grade III, and 5 disks as grade V. In the 12 disks classified as Pfirrmann grade V by the first observer, the second observer agreed in all but 1 disk, which was classified as grade IV instead. When analyzing the differences between the observers according to the Suzuki classification, the first observer rated a single disk as grade 0, which was classified as grade I by the second observer. In total, 29 disks were classified as grade I according to the Suzuki classification by the first observer. Both observers agreed on grade I in 18 cases, while the other disks were classified as grade 0 in 2 cases, grade II in 7 cases, and grade III in 2 cases. For Suzuki grade II, the observers agreed only on 17 out of 50 disks. The other disks classified as grade II by the first observer were classified as grade 0 in 3 cases, as grade I in 23 cases, and grade III in 16 cases. For those 31 disks classified as grade III by the first observer, both observers had the same rating for 25 disks, while 2 were classified as grade I and 4 as grade II.

Discussion

The Suzuki classification was introduced to provide an MRI-based classification tailored for the evaluation of cervical disk degeneration. However, data from this study suggests that it has no advantage over the already well-established Pfirrmann classification regarding inter- or intraobserver reliability. This is particularly remarkable as the Suzuki classification system distinguishes only between four different stages of disk degeneration compared to the Pfirrmann classification, which uses five stages.

The κ-coefficients for intraobserver (κ=0.96) and interobserver (κ=0.90) agreement reported by Suzuki et al.[13] were much higher than in this present study. Pfirrmann et al.[7] reported an intraobserver reliability of κ equal to 0.84-0.90 and interobserver reliability of κ equal to 0.69-0.81 among the three readers. Although the intraobserver reliability was similar to the findings in this study, interrater reliability was higher when compared to this study. Griffith et al.[6] introduced a modified version of the Pfirrmann classification distinguishing eight different stages of disk degeneration, which was subsequently validated for the cervical spine[12] with a near perfect interrater (κ=0.82) and intrarater (κ=0.83-0.92) reliability.

While the intraobserver reliability provided by this study is very comparable to those reported by prior studies, the interrater reliability was substantially lower than in other studies. Analyzing the differences of the rating by the first and second observer, the second observer ascribes less degeneration (a lower Pfirrmann and Suzuki grade) to the analyzed cervical levels. These findings may originate from the classification itself. In the Pfirrmann classification, grade I is defined as a homogeneous, bright white disk. Grade II is defined as a disk that is inhomogeneous and with or without horizontal bands. Grade III is defined as a disk that is inhomogeneously gray. In the Suzuki classification, grade 1 is defined as a homogenous nucleus with high intensity resembling Pfirrmann grade I. Suzuki grade II is defined as an inhomogeneous nucleus that resembles Pfirrmann grades II and III, whereas the Pfirrmann classification does not take disk bulging into account. These definitions may leave some freedom of interpretation and probably led to the relevant difference between the two raters. The slightly better interobserver reliability of the Suzuki classification can be ascribed to the lower number of grades especially in those where the biggest discrepancy was found. Pfirrmann grade V provides an excellent agreement between the two observers, while for Suzuki grade III, there was a greater disagreement between the two observers. The Suzuki classification, however, showed a more even distribution of cervical disk degeneration between grades I, II, and III (Table 2), which is more like to adequately represent the gradual process of disk degeneration. In the Pfirrmann classification, however, most cases were allocated to grades III and IV (Table 1), which can be considered a major disadvantage. Another advantage of the Suzuki classification is its inclusion of cervical disk bulging into the classification. While both classifications consider disk height and signal intensity of the nucleus, usually, both factors have no clinical consequence regarding treatment. However, disk bulging may lead to nerve root or spinal cord compression and in further consequence to the necessity of surgery.

A possible limitation of this study is that we have only analyzed three different levels of the cervical spine. We deemed it unnecessary to analyze all cervical spine levels as previous studies show a highly comparable inter- and intraobserver reliability between all levels of the cervical spine[12]. The selected spinal levels in this study, alongside the age distribution of the evaluated patient cohort, may explain the few cases of low-grade disk degeneration (as per our findings), as the cervical levels C4/5 and C5/6 are among the most commonly degenerated levels of the cervical spine[13].

Despite being established for many years, the Pfirrmann classification, as well as the Suzuki classification, still does not provide information that is necessary for the decision-making in surgery and is therefore not routinely assessed in clinical practice. Therefore, if a new classification should be established, aside from solely and morphologically describing disk degeneration, it should focus on factors that may guide the treatment of patients. For scientific purposes, probably deep learning or machine learning algorithms[17] will supersede the evaluation of spinal degeneration in the near future.

In conclusion, despite being specifically designed for the evaluation of cervical disk degeneration, there seems to be no advantage of grading cervical spine degeneration using the Suzuki classification rather than the well-established Pfirrmann classification in terms of inter- or intraobserver reliability. While the Pfirrmann classification has the advantage of being better established, the Suzuki classification may be clinically superior due to a better representation of cervical disk degeneration and the consideration of disc bulging for the classification of cervical disc degeneration.

Conflicts of Interest: The authors declare that there are no relevant conflicts of interest.

Author Contributions: Urbanschitz Lukas; Conception and design, analysis and interpretation of data, and creation of the manuscript.

Susanne Bensler; Acquisition of data and revision of the manuscript for critically important intellectual content.

Sascha Merat; Acquisition of data and revision of the manuscript for critically important intellectual content.

Christopher G. Lenz; Analysis and interpretation of data and revision of the manuscript for critically important intellectual content.

Karim Eid; Project Supervisor and Co-creator of the manuscript

Ethical Approval: 2020-01001 (Swissethics)

References

1. Matsumoto M, Okada E, Ichihara D, et al. Age-related changes of thoracic and cervical intervertebral discs in asymptomatic subjects. Spine (Phila Pa 1976). 2010;35(14):1359-64.
2. Matsumoto M, Okada E, Toyama Y, et al. Tandem age-related lumbar and cervical intervertebral disc changes in asymptomatic subjects. Eur Spine J. 2013;22(4):708-13.
3. Nouri A, Tetreault L, Singh A, et al. Degenerative cervical myelopathy: Epidemiology, genetics, and pathogenesis. Spine. 2015;40(12):E675-93.
4. Peng B, DePalma MJ. Cervical disc degeneration and neck pain. J Pain Res. 2018;11:2853-7.
5. Boos N, Weissbach S, Rohrbach H, et al. Classification of age-related changes in lumbar intervertebral discs: 2002 Volvo Award in basic science. Spine (Phila Pa 1976). 2002;27(23):2631-44.
6. Griffith JF, Wang YXJ, Antonio GE, et al. Modified Pfirrmann grading system for lumbar intervertebral disc degeneration. Spine (Phila Pa 1976). 2007;32(24):E708-12.
7. Pfirrmann CWA, Metzdorf A, Zanetti M, et al. Magnetic resonance classification of lumbar intervertebral disc degeneration. Spine (Phila Pa 1976). 2001;26(17):1873-8.
8. Oh CH, Kim DY, Ji GY, et al. Cervical arthroplasty for moderate to severe disc degeneration: Clinical and radiological assessments after a minimum follow-up of 18 months -pfirrmann grade and cervical arthroplasty-. Yonsei Med J. 2014;55(4):1072-9.
9. Nakashima H, Yukawa Y, Suda K, et al. Cervical disc protrusion correlates with the severity of cervical disc degeneration: A cross-sectional study of 1211 relatively healthy volunteers. Spine (Phila Pa 1976). 2015;40(13):E774-9.

10. Cuellar VG, Cuellar JM, Vaccaro AR, et al. Accelerated degeneration after failed cervical and lumbar nucleoplasty. J Spinal Disord Tech. 2010;23(8):521-4.

11. Markotić V, Zubac D, Miljko M, et al. Level of education as a risk factor for extensive prevalence of cervical intervertebral disc degenerative changes and chronic neck pain. Cent Eur J Public Health. 2017;25(3):245-50.

12. Xu YZ, Wang YT, Fan P, et al. Complications and outcomes of open posterior lumbar spinal fusion surgery in obese patients: A meta-analysis. Br J Neurosurg. 2020:1-9.

13. Suzuki A, Daubs MD, Hayashi T, et al. Magnetic resonance classification system of cervical intervertebral disk degeneration: Its Validity and Meaning. Clin Spine Surg. 2017;30(5):E547-53.

14. Suzuki A, Daubs MD, Hayashi T, et al. Patterns of cervical disc degeneration: Analysis of magnetic resonance imaging of over 1000 symptomatic subjects. Glob Spine J. 2018;8(3):254-9.

15. Habibi H, Suzuki A, Tamai K, et al. The Severity of cervical disc degeneration does not impact 2-year postoperative outcomes in patients with cervical spondylotic myelopathy who underwent laminoplasty. Spine (Phila Pa 1976). 2020;45(18):E1142-9.

16. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159-74.

17. Huang J, Shen H, Wu J, et al. Spine Explorer: A deep learning based fully automated program for efficient and reliable quantifications of the vertebrae and discs on sagittal lumbar spine MR images. Spine J. 2020;20(4):590-9.