Qualitative and quantitative assessment of degeneration of cervical intervertebral discs and facet joints

Joris Walraevens · Baoge Liu · Jos Vander Sloten · Jan Goffin

Received: 10 March 2008 / Revised: 14 August 2008 / Accepted: 17 October 2008 / Published online: 13 November 2008
© Springer-Verlag 2008

Abstract Degeneration of intervertebral discs and facet joints is one of the most frequently encountered spinal disorders. In order to describe and quantify degeneration and evaluate a possible relationship between degeneration and biomechanical parameters, e.g., the intervertebral range of motion and intradiscal pressure, a scoring system for degeneration is mandatory. However, few scoring systems for the assessment of degeneration of the cervical spine exist. Therefore, two separate objective scoring systems to qualitatively and quantitatively assess the degree of cervical intervertebral disc and facet joint degeneration were developed and validated. The scoring system for cervical disc degeneration consists of three variables which are individually scored on neutral lateral radiographs: “height loss” (0–4 points), “anterior osteophytes” (0–3 points) and “endplate sclerosis” (0–2 points). The scoring system for facet joint degeneration consists of four variables which are individually scored on neutral computed tomography scans: “hypertrophy” (0–2 points), “osteophytes” (0–1 point), “irregularity” on the articular surface (0–1 point) and “joint space narrowing” (0–1 point). Each variable contributes with varying importance to the overall degeneration score (max 9 points for the scoring system of cervical disc degeneration and max 5 points for facet joint degeneration). Degeneration of 20 discs and facet joints of 20 patients was blindly assessed by four raters: two

neurosurgeons (one senior and one junior) and two radiologists (one senior and one junior), firstly based on first subjective impression and secondly using the scoring systems. Measurement errors and inter- and intra-rater agreement were determined. The measurement error of the scoring system for cervical disc degeneration was 11.1 versus 17.9% of the subjective impression results. This scoring system showed excellent intra-rater agreement (ICC = 0.86, 0.75–0.93) and excellent inter-rater agreement (ICC = 0.78, 0.64–0.88). Surgeons as well as radiologists and seniors as well as juniors obtained excellent inter- and intra-rater agreement. The measurement error of the scoring system for cervical facet joint degeneration was 20.1 versus 24.2% of the subjective impression results. This scoring system showed good intra-rater agreement (ICC = 0.71, 0.42–0.89) and fair inter-rater agreement (ICC = 0.49, 0.26–0.74). Both scoring systems fulfilled the criteria for recommendation proposed by Kettler and Wilke. Our scoring systems can be reliable and objective tools for assessing cervical disc and facet joint degeneration. Moreover, the scoring system of cervical disc degeneration was shown to be experience- and discipline-independent.

Keywords Disc and facet joint degeneration · Scoring system · Cervical spine

J. Walraevens (✉) · J. Vander Sloten
Division of Biomechanics and Engineering Design,
KULeuven, Celestijnenlaan 300C, PB 2419,
3001 Heverlee, Belgium
e-mail: joris.walraevens@mech.kuleuven.be

B. Liu · J. Goffin
Division of Experimental Neurosurgery and Neuroanatomy,
KULeuven, University Hospital Gasthuisberg, Leuven, Belgium

Background

Degeneration of intervertebral discs and facet joints is one of the most frequently encountered spinal disorders [20]. In order to describe and quantify degeneration and evaluate a possible relationship between degeneration and biomechanical parameters, e.g., the intervertebral range of

motion, sagittal alignment and intradiscal pressure, a scoring system is mandatory. Moreover, a scoring system can be a helpful tool to investigate the possible correlation between intervertebral disc degeneration and facet joint degeneration or to assess the evolution of degeneration over time after an arthrodesis or after arthroplasty.

However, as of date a limited number of scoring systems for degeneration of cervical intervertebral discs and facet joints based on radiographs have been developed [12]. Two scoring systems for cervical disc degeneration have been tested for reliability (Kellgren et al. [10] by Côté et al. [4] and Kettler et al. [11]). Only one scoring system for facet joint degeneration has been tested for reliability (Kellgren et al. [10]).

In their review Kettler and Wilke observed a wide variety in design and terminology of the existing scoring systems [12]. One of the major drawbacks of the scoring systems of Kellgren for cervical disc and facet joint degeneration is the use of subjective, descriptive terms as “moderate”, “severe” to quantify degeneration. To ensure objectivity, Kettler et al. [11] developed a numerical radiographic scoring system for cervical intervertebral disc degeneration. In this scoring system three variables: “height loss” of intervertebral disc height, “osteophyte formation” and “diffuse sclerosis” have to be graded individually on a scale from 0 to 3. Based on the sum, the overall degree of disc degeneration is determined. Although Kettler and Wilke obtained substantial inter-rater agreement ($\kappa = 0.688$), the scoring system has some drawbacks. Firstly, the scoring system is difficultly applicable in daily clinical practice because it is complex and time-consuming. Secondly, the scoring system was developed based on the lateral radiographs of human cadaveric osteoligamentous spine specimens. It has not been tested in vivo.

A reliable scoring system for the assessment of cervical facet joint degeneration does not exist up to date. Kellgren

et al. [10] used lateral radiographs to score the degeneration of cervical facet joints. Côté et al. [4] found an unacceptable inter-rater agreement for this scoring system (ICC = 0.45). They claimed that one of the reasons for this poor agreement was that lateral radiographs often poorly visualize the facet joints. Computed tomography scans might improve the visibility and therefore early degenerative change might be better detected.

The goal of this study is to establish and validate a quantitative scoring system for cervical intervertebral disc degeneration based on lateral radiographs and a scoring system for cervical facet joint degeneration based on computed tomography scans. The results of these scoring systems are compared with results based on first subjective impression. Moreover, as an application of both scoring systems, the spatial correlation between facet joint degeneration and intervertebral disc degeneration is investigated.

Materials and methods

Scoring system for cervical disc degeneration based on lateral radiographs

The scoring system for cervical disc degeneration consists of three variables with decreasing importance to the total degeneration score: “height loss”, “anterior osteophytes”, and “endplate sclerosis”. Each of these variables is individually scored. Next, the three variables are summed to the overall degree of disc degeneration (ranging from 0 to 9; Table 1). Height loss is defined as the middle disc height with respect to a normal middle disc height at an adjacent level. Height loss is graded from 0 to 4. Middle disc height of the target level is assessed with respect to the middle

Table 1 Scoring system of cervical disc degeneration based on neutral lateral radiographs

1. Height loss		
Middle disc height compared to normal middle disc height at an adjacent level	0%	0 points
	≤25%	1 points
	>25%–≤50%	2 points
	>50%–≤75%	3 points
	>75%	4 points
2. Anterior osteophytes with respect to the AP diameter of the corresponding VB	No osteophytes	0 points
	≤1/8 AP diameter	1 point
	>1/8–≤1/4 AP diameter	2 points
	>1/4 AP diameter	3 points
3. Endplate sclerosis	No sclerosis	0 points
	Detectable	1 point
	Definite	2 points
Overall degree of disc degeneration = 1 + 2 + 3	0 points (no degeneration)	
	1–3 points (mild degeneration)	
	4–6 points (moderate degeneration)	
	7–9 points (severe degeneration)	

VB vertebral body,
AP anteroposterior

Fig. 1 Height loss is assessed on lateral radiographs. Middle disc height of the target level (*CD*) is compared to the middle disc height of a normal adjacent level (*AB*). No height loss is scored as 0; a loss in disc height less than 25% receives 1 point; height loss between 25% and 50% receives 2 points; between 50% and 75% 3 points; and 4 points are given when the height loss is more than 75%

disc height of a normal adjacent level (Fig. 1). The length of the anterior osteophytes is measured with respect to the anteroposterior diameter of the corresponding vertebral body, which is measured at the middle of the vertebral body (Fig. 2). Anterior osteophytes are scored from 0 to 3. When different scores are attributed to the cranial and caudal anterior corners of the target level, the highest score is chosen. For endplate sclerosis, a distinction between no apparent sclerosis, just detectable and definite sclerosis is made (score 0–2, Fig. 3).

Scoring system for cervical facet joint degeneration based on computed tomography scans

The scoring system for degeneration of the facet joints consists of four variables with varying importance to the total score: “hypertrophy”, “osteophytes”, “irregularity” of the articular surface, and “joint space narrowing”. These variables are assessed on computed tomography scans. Similarly to the scoring system for cervical disc degeneration, each of these variables is individually scored. Next, the four

Fig. 2 Anterior osteophytes are assessed on lateral radiographs. The length of the anterior osteophytes (*XY* and *PQ*) is measured with respect to the anteroposterior diameter of the vertebral body (*AB* and *CD*, respectively). When no anterior osteophyte is visible, a score of 0 is attributed; an anterior osteophyte that is just detectable receives 1 point; an anterior osteophyte which extends less than one-fourth of the anteroposterior diameter receives 2 points; when the anterior osteophytes extends more than one fourth, 3 points are attributed. The highest of the cranial and caudal score is used as final score

variables are summed to obtain the overall degree of facet joint degeneration (ranging from 0 to 5; see Table 2).

Hypertrophy is graded from 0 to 2. Zero points are given when there is no hypertrophy present; 1 point when hypertrophy is visible on one of the margins of the articular surface; and 2 points when it is present on all margins (Fig. 4). Osteophytes are graded 0 if no osteophytes are present and 1 if osteophytes are present (Fig. 5). Irregularity of the articular surface is scored 0 if the articular surface is smooth; and is scored 1 if the articular surface is irregular (Fig. 6). If the joint space of both facet joints is not narrowed, joint space narrowing is scored 0. In case of narrowing, it is scored 1 (Fig. 7). If a difference in degeneration score between the left and right facet joint is found, the highest of both scores is used.

Experimental procedure

In a retrospective study, neutral lateral radiographs and computed tomography scans of 20 patients, recently

Fig. 3 Endplate sclerosis is assessed on lateral radiographs. Zero points are attributed if no endplate sclerosis is present (a); 1 point is given if sclerosis is just detectable (b); 2 points are given when sclerosis is definitively present at the cranial and/or the caudal endplate (c)

Table 2 Scoring system of cervical facet joint degeneration based on computed tomography scans

1. Hypertrophy of FJ	None	0 points
	On one of the margins of the articular surfaces	1 point
	On all margins of the articular surfaces	2 points
2. Osteophytes on FJ	None	0 points
	Yes	1 point
3. Irregularity on articular surface	Normal	0 points
	Irregular	1 point
4. Joint space narrowing	Normal	0 points
	Narrowed	1 point
Overall degree of facet joint degeneration = 1 + 2 + 3 + 4	0 points (no degeneration)	
	1 point (mild degeneration)	
	2–3 points (moderate degeneration)	
	4–5 points (severe degeneration)	

FJ facet joint

Fig. 4 Hypertrophy is assessed on transverse computed tomography scans. Zero points are given when no hypertrophy is present (a); 1 point when hypertrophy is present on one of the margins of the articular surface (b); and 2 points when it is present on all margins (c)

operated for cervical degenerative disc disease, were used for the assessment of intervertebral disc and facet joint degeneration. The name of each patient was removed and the clinical history remained unknown to prevent bias.

Twenty intervertebral discs and facet joints of the operated level were analyzed by four raters: two neurosurgeons (one senior and one junior) and two radiologists (one senior and one junior). None of the raters were previously connected to the study. Written instructions were provided to all raters before the assessment. No assistance was given during the assessment.

Intervertebral disc degeneration and facet joint degeneration were assessed three times: firstly based on subjective impression (SI; see Table 3), followed by a second time using the scoring systems (SS1). After 1 month, the raters were asked to reassess all levels a third time using the scoring systems (SS2). Between all assessments, the order was randomized to prevent bias.

Intervertebral disc degeneration was scored based on neutral lateral radiographs; facet joint degeneration for the operated level was assessed on computed tomography scans.

Fig. 5 Osteophytes are assessed on transverse computed tomography scans. Zero points are given when no osteophytes on either of the facet joints are visible (**a**); 1 point when osteophytes are present (**b**)

Fig. 6 Irregularity of the articular surface is assessed on transverse computed tomography scans. Zero points are given when the articular surface of either of the facet joints is smooth (**a**); 1 point when the surface is irregular (**b**)

Fig. 7 Joint space narrowing is assessed on computed tomography scans. Zero points are given when the joint space of either of the facet joints is not narrowed (**a**); 1 point when the space is narrowed (**b**)

Table 3 Overall degree of degeneration of cervical disc and facet joints based on first subjective impression

Indication
No degeneration
Mild degeneration
Moderate degeneration
Severe degeneration

Statistical analysis

The measurement error was estimated using within-subject standard deviations based on the SI and SS1 results. The measurement error is a measure for the variation in the scoring system [1, 2]. Ninety-five percent prediction limits

can be calculated using the measurement error. The difference between the observed value and the measured value is expected to be less than this value in 95% of the observations.

Inter-rater agreement, i.e., the agreement between the ratings of all raters, and the intra-rater agreement, i.e., the agreement between the ratings of the same rater, were evaluated using a two-way random model of intra-class correlation coefficients (ICC), with measures of absolute agreement [18]. A single-measure intra-class correlation was selected to estimate the reliability of a single rating instead of a mean of several ratings. Inter-rater agreement was assessed based on the SI and SS1 results. Intra-rater agreement was calculated based on the SS1 and SS2 results. Ninety-five percent confidence intervals (CI) were

Table 4 Convention for inter- and intra-rater agreement according to Fleiss et al. [7] ICC: Intra-class correlation coefficient

ICC	Strength of agreement
$ICC \leq 0.40$	Poor agreement
$0.40 < ICC \leq 0.60$	Fair agreement
$0.60 < ICC \leq 0.75$	Good agreement
$0.75 < ICC$	Excellent agreement

constructed around each ICC [22]. Table 4 provides the convention that is used throughout the text.

Linear correlations were investigated using Pearson r correlation coefficients. Data analysis was performed using Statistica 6.0.

Results

Disc degeneration

Measurement error

The scoring system shows an improved measurement error for the overall degree of disc degeneration the with regard to the SI result (11.1 vs. 17.9%; Table 5). The variable “endplate sclerosis” has the largest measurement error (33.9%).

Inter-rater agreement

The inter-rater agreement for the overall degree of disc degeneration of the scoring system is excellent

(ICC = 0.78, 0.64–0.88; Table 6). The ICC of the variable “endplate sclerosis” is poor (ICC = 0.31 vs. 0.73 for height loss and anterior osteophytes). The overall degree of disc degeneration shows a small improvement in ICC with respect to the SI results (ICC = 0.77, 0.59–0.89).

Intra-rater agreement

Excellent intra-rater agreement is observed for the overall degree of disc degeneration (ICC = 0.86, 0.75–0.93; Table 7). This observation holds for all variables individually, except for endplate sclerosis which has a good intra-rater reliability (ICC = 0.62, 0.40–0.80).

Comparison between raters with different experience

A comparison between experienced and inexperienced raters was made. Senior raters obtained better inter-rater agreement than junior raters for the variables height loss and endplate sclerosis (Fig. 8). Junior raters showed better inter-rater agreement for the SI results and anterior osteophytes. For all variables, except for endplate sclerosis, junior as well as senior raters obtained excellent intra-rater agreement (Fig. 9), with best results for the junior raters.

Junior as well as senior raters obtained good inter-rater and excellent intra-rater agreement for the overall degree of disc degeneration ($p > 0.05$).

Comparison between raters of different disciplines

A comparison between raters of different disciplines was made: one senior and one junior surgeon versus one senior

Table 5 Measurement error of cervical disc degeneration based on the assessment of 20 intervertebral discs [relative within-subject standard deviation (WSSD) and 95% prediction limit (PL)]

	Scale	WSSD (%)	95% PL
Subjective impression of the overall degree of disc degeneration		17.87	1.0509
Height loss	0–4	12.29	0.9635
Anterior osteophytes	0–3	17.21	1.0121
Endplate sclerosis	0–2	33.85	1.3269
Overall degree of disc degeneration	0–9	11.11	1.9600

Table 6 Inter-rater agreement between all raters based on the assessment of 20 intervertebral discs [Intra-class correlation coefficients (ICC) and 95% confidence intervals (CI)]

	ICC	95% CI
Subjective impression of the overall degree of disc degeneration	0.7650	0.5913–0.8874
Height loss	0.7284	0.5403–0.8673
Anterior osteophytes	0.7275	0.5813–0.8586
Endplate sclerosis	0.3107	0.1273–0.5582
Overall degree of disc degeneration	0.7759	0.6421–0.8871

Table 7 Intra-rater agreement between two assessments (SS1 and SS2) of all raters based on the assessment of 20 intervertebral discs [Intra-class correlation coefficients (ICC) and 95% confidence intervals (CI)]

	ICC	95% CI
Height loss	0.8039	0.6036–0.9137
Anterior osteophytes	0.7869	0.5847–0.9034
Endplate sclerosis	0.6156	0.3978–0.8010
Overall degree of disc degeneration	0.8580	0.7461–0.9338

and one junior radiologist. The radiologists obtained better inter-rater agreement for the overall degree of disc degeneration ($p < 0.05$). The inter-rater agreement of all variables was better for the radiologists, except for anterior osteophytes. The intra-rater agreement of all variables was better for the radiologists, except for height loss.

Surgeons as well as radiologists obtained excellent inter-rater and intra-rater agreement for the overall degree of disc degeneration (Figs. 8, 9).

Facet joint degeneration

Measurement error

The overall degree of facet joint degeneration shows an improved measurement error in comparison to the SI result (20.1 vs. 24.2%; Table 8). Nevertheless, the measurement

errors remain large for the overall degree of facet joint degeneration (20.1%) and all of the variables individually (39.3–43.6%).

Inter-rater agreement

Poor inter-rater agreement for the SI results (ICC = 0.35, 0.16–0.59; Table 9) and fair inter-rater agreement of the overall degree of facet joint degeneration (ICC = 0.49, 0.26–0.74) is obtained. Joint space narrowing was the variable with the highest inter-rater agreement (ICC = 0.40 compared to 0.17, 0.39 and 0.20 for hyper-trophy, osteophytes and irregularity).

Intra-rater agreement

Good intra-rater agreement was obtained for the overall degree of facet joint degeneration (ICC = 0.72, 0.42–0.89). However, for the variables individually, with the exception of osteophytes, fair intra-rater agreement was found (Table 10).

Comparison between raters with different experience

A comparison between experienced and inexperienced raters was made. Senior raters obtained better inter-rater agreement than junior raters for the SI results and the overall degree of facet joint degeneration (Fig. 10).

Fig. 8 Inter-rater agreement between senior versus junior raters and surgeons versus radiologists and between all raters based on the assessment of 20 intervertebral discs (Intra-class correlation coefficients (ICC); * $p < 0.05$)

Fig. 9 Intra-rater agreement between two assessments (SS1 and SS2) of senior versus junior raters and surgeons versus radiologists and of all raters based on the assessment of 20 intervertebral discs (Intra-class correlation coefficients (ICC); * $p < 0.05$)

Table 8 Measurement error of cervical facet joint degeneration based on the assessment of 20 facet joints [relative within-subject standard deviation (WSSD) and 95% prediction limit (PL)]

	Scale	WSSD (%)	95% PL
Subjective impression of the overall degree of facet joint degeneration		24.25	1.4258
Hypertrophy	0–2	30.37	1.1907
Osteophytes	0–1	40.09	0.7857
Irregularities	0–1	43.64	0.8554
Joint space narrowing	0–1	39.34	0.7711
Overall degree of facet joint degeneration	0–5	20.06	1.9658

Table 9 Inter-rater agreement between all raters based on the assessment of 20 facet joints [Intra-class correlation coefficients (ICC) and 95% confidence intervals (CI)]

	ICC	95% CI
Subjective impression of the overall degree of facet joint degeneration	0.3494	0.1647–0.5902
Hypertrophy	0.1708	0.0000–0.4759
Osteophytes	0.3979	0.1707–0.6817
Irregularities	0.2031	0.0128–0.5139
Joint space narrowing	0.4007	0.1769–0.6837
Overall degree of facet joint degeneration	0.4866	0.2589–0.7449

Senior raters obtained excellent intra-rater agreement for the overall degree of facet joint degeneration, juniors obtained good agreement ($p < 0.05$; Fig. 11).

Comparison between raters of different disciplines

A comparison between raters of different disciplines was made: one senior and one junior surgeon versus one

senior and one junior radiologist. Surgeons obtained better inter-rater agreement than radiologists for the overall degree of facet joint degeneration and all of its variables, except for hypertrophy (Fig. 10). The surgeons obtained good inter-rater and excellent intra-rater agreement for the overall degree of facet joint degeneration and the radiologists fair and good agreement, respectively (Fig. 11).

Table 10 Intra-rater agreement between two assessments (SS1 and SS2) of all raters based on the assessments of 20 facet joints [Intra-class correlation coefficients with 95% confidence intervals (CI)]

	ICC	95% CI
Hypertrophy	0.5596	0.2400–0.8161
Osteophytes	0.6568	0.3818–0.8576
Irregularities	0.5443	0.2106–0.8040
Joint space narrowing	0.5176	0.2067–0.7902
Overall degree of facet joint degeneration	0.7167	0.4230–0.8894

Correlation between cervical disc and facet joint degeneration

As shown in Table 11, a significant but weak correlation is observed between disc and facet joint degeneration based on the SI results (Pearson r : 0.33, $p < 0.05$) as well as based on total degeneration scores of disc and facet joint degeneration (Pearson r : 0.27, $p < 0.05$).

Discussion

In this study separate scoring systems for cervical intervertebral disc and facet joint degeneration were proposed and tested for inter- and intra-rater agreement. Using these scoring systems, the spatial correlation between disc and facet joint degeneration was assessed.

Only two scoring systems for the assessment of cervical intervertebral disc degeneration have been tested for inter-

rater agreement up to date: Kellgren et al. [10] by Côté et al. [4] and Kettler et al. [11]. Both scoring systems obtained good inter-rater agreement (ICC = 0.71 and $\kappa = 0.688$), compared to the excellent inter-rater agreement of our scoring system (ICC = 0.78).

Our scoring system for cervical intervertebral disc degeneration has some similarities with the scoring system of Kettler et al. Three variables have to be graded individually on a numerical scale based on objective criteria; the sum of these scores assigns the overall degree of degeneration. Nevertheless, our scoring system is fundamentally different. In contrast to the scoring system of Kettler et al., only middle disc height is used to calculate height loss. Moreover, only anterior osteophytes, but no posterior osteophytes, are assessed in our scoring system. A third difference is that the variables contribute to the overall degree of degeneration with variable importance. Height loss of the middle disc height has the highest importance (four points on a nine-point scale), followed by anterior osteophytes and endplate sclerosis (three and two points on a nine-point scale). This strategy was chosen because height loss is a straightforward and accurate indicator for disc degeneration. In contrast to posterior osteophytes, which are often not clearly visible on lateral radiographs due to the overlap with the lateral processes, anterior osteophytes are easily scored. Anterior osteophytes have however less clinical importance and therefore lower importance than for height loss was assigned in our scoring system. Endplate sclerosis contributes to disc degeneration, but the scoring of endplate sclerosis is very sensitive to the

Fig. 10 Inter-rater agreement between senior versus junior raters and surgeons versus radiologists and between all raters based on the assessment of 20 facet joints (Intra-class correlation coefficients (ICC); * $p < 0.05$)

Fig. 11 Intra-rater agreement between two assessments (SS1 and SS2) of senior versus junior raters and surgeons versus radiologists and of all raters based on the assessment of 20 facet joints (Intra-class correlation coefficients (ICC); * $p < 0.05$)

Table 11 Correlation table of Pearson r coefficients of intervertebral disc degeneration versus facet joint degeneration based on the subjective impression (SI) results and the scoring systems. Significant correlations ($p < 0.05$) are displayed in **bold**

	SI disc	Height loss	Anterior osteophytes	Endplate sclerosis	Overall degeneration (disc)
SI facet joint	0.33	0.22	0.04	0.10	0.19
Hypertrophy	0.36	0.26	0.08	0.20	0.27
Osteophytes	0.45	0.38	0.10	0.21	0.35
Irregularities	0.14	0.23	−0.09	−0.11	0.03
Joint space narrowing	−0.01	0.02	−0.07	0.07	0.01
Overall degeneration (facet)	0.38	0.36	0.02	0.16	0.27

quality of the radiographs and the proper alignment of the intervertebral disc. Therefore, the lowest importance was attributed to endplate sclerosis.

These modifications did not lead to a lower inter-rater agreement. On the contrary, a stronger inter-rater agreement was observed. According to Kettler and Wilke, the inter-rater reliability of the scoring system fulfills their criterion for recommendation ($ICC > 0.60$) [12].

In addition to the inter-rater agreement, also intra-rater agreement of the scoring system for intervertebral disc degeneration was calculated. This value is a measure for the reproducibility of the scoring system. Excellent intra-rater agreement was observed for our scoring system.

A drawback of this study is that no validation of the scoring system against a gold standard was performed. The excellent inter- and intra-rater agreement indicates that the scoring system is highly reliable and repeatable. However, such agreement does not eliminate the possibility

of a systematic error (consistent over- or underestimation of the “real” degree of degeneration). Kettler and Wilke validated their scoring system based on lateral radiographs of human cadaveric osteoligamentous spine specimens, and used macroscopic slices of the respective cadaveric specimens to assess the “real” degree of degeneration. They found that the “real” degree of disc degeneration was underestimated in 64% of all discs [11]. However, the use of the cadaveric specimens might have influenced the results. The surrounding soft tissues that can decrease visibility of the intervertebral disc space are removed. And in contrast to in vivo measurements, a long exposure time can be used. This increases contrast on the lateral radiographs, providing a better visibility.

Next to the cadaveric specimens, MRI might have been used as a comparative method to further validate our scoring system. Several scoring systems have been developed to assess cervical intervertebral disc degeneration

based on MRI [5, 13, 16]. Miyazaki et al. [13] claimed that MRI is the most sensitive method for the clinical assessment of intervertebral disc pathology. They reported excellent intra-rater reliability and good to excellent inter-rater reliability for their scoring system. Similar to this study, Miyazaki et al. limited the validation of their scoring system to inter- and intra-rater reliability testing. No comparison against a gold standard, such as cadaveric specimens, was made. Four spinal surgeons acted as observers in their study; no information on their experience was given. In our study, both surgeons and radiologists, juniors as well as seniors acted as observers, illustrating the multi-experience and multi-discipline use of our scoring system. Christe et al. reported that both radiographs and MRI significantly, but weakly, correlated with histology ($r = 0.33$ and $r = 0.49$, $p < 0.05$) in the detection of pathologic lesions in the cervical spine [5]. They did, however, not report on the correlation between MRI and planar radiographs.

As this is a retrospective study, no MRI was available for all patients. A comparison of our scoring system with MRI could therefore not be made.

As this scoring system uses standard lateral radiographs, and as it requires uncomplicated input for the user, the scoring system can easily be used in daily clinical practice for the assessment of cervical disc degeneration.

Senior and junior raters obtained good inter-rater agreement and excellent intra-rater agreement. Surgeons and radiologists obtained excellent inter- and intra-rater agreement. These results indicate that the scoring system can be reliably used by both experienced as well as inexperienced raters from different disciplines.

Only one scoring system for facet joint degeneration has previously been tested for reliability (Kellgren et al. [10] by Côté et al. [4]). Fair inter-rater agreement was found ($ICC = 0.45$). In contrast to Côté et al. [4], who believed that this level of agreement is not acceptable for rigorous outcomes research, Kettler and Wilke [12] noted that it fulfilled their criteria for recommendation ($ICC > 0.40$). According to this criterion, also our scoring system for the assessment of facet joint degeneration ($ICC = 0.49$), can be recommended.

Our scoring system assesses the presence of hypertrophy, osteophytes, irregularities on the articular surface and joint space narrowing at the target level based on computed tomography scans.

Similar to the scoring system of cervical disc degeneration, a drawback of this study is that the scoring system for cervical facet joint degeneration is not compared with a gold standard, such as cadaveric specimens, nor is it compared with an alternative method, such as MRI.

As a clinical application, this scoring system is very useful when degeneration of one patient has to be assessed

and compared at different time intervals, e.g., to investigate the influence of an arthrodesis or arthroplasty on the degeneration of the levels adjacent to the treated level. Moreover, this scoring system is applicable when degeneration of different patients on a certain time point has to be compared. In these cases of relative comparison, a possible systematic error is canceled out.

As an additional application for both scoring systems, the spatial correlation between intervertebral disc and facet joint degeneration has been investigated. In contrast to the lumbar spine [3, 6, 8, 9, 14, 15, 17, 19, 21, 23], this correlation has not been thoroughly investigated for the cervical spine. A weak but significant spatial correlation between cervical intervertebral disc degeneration and facet joint degeneration was observed. However, as this was not a follow-up study, the temporal correlation could not be identified. Therefore, the hypothesis that “disc degeneration precedes facet joint osteoarthritis” [3, 8, 23] cannot be confirmed, nor denied.

Conclusion

Our scoring system for cervical disc degeneration can be a reliable and objective tool. Moreover, this scoring system showed to be experience- and discipline-independent.

Our scoring system for facet joint degeneration, which is based on computed tomography scans, is less reliable. Nevertheless, it fulfills the criteria for recommendation proposed by Kettler and Wilke.

A weak spatial correlation between cervical intervertebral disc and facet joint degeneration has been observed.

Acknowledgments Medtronic Sofamor Danek Inc., USA, is acknowledged for the institutional research and education grant.

References

1. Bland JM, Altman DG (1996) Measurement error and correlation coefficients. *BMJ* 313(7048):41–42
2. Bland JM, Altman DG (1996) Measurement error proportional to the mean. *BMJ* 313:7049–7106
3. Butler D, Trafimow JH, Andersson GB, McNeill TW, Huckman MS (1990) Discs degenerate before facets. *Spine* 15(2):111–113. doi:10.1097/00007632-199002000-00012
4. Côté P, Cassidy JD, Yong-Hing K, Sibley J, Loewy J (1997) Apophyseal joint degeneration, disc degeneration, and sagittal curve of the cervical spine. Can they be measured reliably on radiographs? *Spine* 22(8):859–864. doi:10.1097/00007632-199704150-00007
5. Christe A, Läubli R, Guzman R, Berlemann U, Moore RJ, Schroth G et al (2005) Degeneration of the cervical disc: histology compared with radiography and magnetic resonance imaging. *Neuroradiology* 47(10):721–729. doi:10.1007/s00234-005-1412-6
6. Derian TC (1994) Adjacent disc degeneration in patients with prior spinal fusion procedures. *Spine* 19(19):2244–2245. doi:10.1097/00007632-199410000-00024

7. Fleiss J (1986) Reliability of measurement. In: The design and analysis of clinical experiments. Wiley, New York, pp 1–32
8. Fujiwara A, Lim TH, An HS, Tanaka N, Jeon CH, Andersson GB et al (2000) The effect of disc degeneration and facet joint osteoarthritis on the segmental flexibility of the lumbar spine. *Spine* 25(23):3036–3044. doi:[10.1097/00007632-200012010-00011](https://doi.org/10.1097/00007632-200012010-00011)
9. Gries NC, Berlemann U, Moore RJ, Vernon-Roberts B (2000) Early histologic changes in lower lumbar discs and facet joints and their correlation. *Eur Spine J* 9(1):23–29. doi:[10.1007/s005860050004](https://doi.org/10.1007/s005860050004)
10. Kellgren J, Jeffrey M, Ball J (1963) The epidemiology of chronic rheumatism, vol. II. Blackwell, Oxford, pp 14–19
11. Kettler A, Rohlmann F, Neidlinger-Wilke C, Werner K, Claes L, Wilke HJ (2006) Validity and interobserver agreement of a new radiographic grading system for intervertebral disc degeneration. Part II. Cervical spine. *Eur Spine J* 15(6):732–741. doi:[10.1007/s00586-005-1037-9](https://doi.org/10.1007/s00586-005-1037-9)
12. Kettler A, Wilke HJ (2006) Review of existing grading systems for cervical or lumbar disc and facet joint degeneration. *Eur Spine J* 15(6):705–718. doi:[10.1007/s00586-005-0954-y](https://doi.org/10.1007/s00586-005-0954-y)
13. Kolstad F, Myhr G, Kvistad KA, Nygaard OP, Leivseth G (2005) Degeneration and height of cervical discs classified from MRI compared with precise height measurements from radiographs. *Eur J Radiol* 55(3):415–420. doi:[10.1016/j.ejrad.2005.02.005](https://doi.org/10.1016/j.ejrad.2005.02.005)
14. Kumaresan S, Yoganandan N, Pintar FA, Maiman DJ, Goel VK (2001) Contribution of disc degeneration to osteophyte formation in the cervical spine: a biomechanical investigation. *J Orthop Res* 19(5):977–984. doi:[10.1016/S0736-0266\(01\)00010-9](https://doi.org/10.1016/S0736-0266(01)00010-9)
15. Lewin T (1964) Osteoarthritis in lumbar synovial joints. a morphologic study. *Acta Orthop Scand Suppl* S73:1–12
16. Miyazaki M, Hong SW, Yoon SH, Morishita Y, Wang JC (2008) Reliability of a magnetic resonance imaging-based grading system for cervical intervertebral disc degeneration. *J Spinal Disord Tech* 21(4):288–292. doi:[10.1097/BSD.0b013e31813c0e59](https://doi.org/10.1097/BSD.0b013e31813c0e59)
17. Moore RJ, Crotti TN, Osti OL, Fraser RD, Vernon-Roberts B (1999) Osteoarthritis of the facet joints resulting from annular rim lesions in sheep lumbar discs. *Spine* 24(6):519–525. doi:[10.1097/00007632-199903150-00003](https://doi.org/10.1097/00007632-199903150-00003)
18. Shrout P, Fleiss J (1979) Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 86(2):420–428. doi:[10.1037/0033-2909.86.2.420](https://doi.org/10.1037/0033-2909.86.2.420)
19. Swanepoel MW, Adams LM, Smeathers JE (1995) Human lumbar apophyseal joint damage and intervertebral disc degeneration. *Ann Rheum Dis* 54(3):182–188. doi:[10.1136/ard.54.3.182](https://doi.org/10.1136/ard.54.3.182)
20. Szpalski M, Gunzburg R, Mayer M (2002) Spine arthroplasty: a historical review. *Eur Spine J* 11(S2):S65–S84
21. Terti M, Paaanen H, Laato M, Aho H, Komu M, Korman M (1991) Disc degeneration in magnetic resonance imaging a comparative biochemical, histologic, and radiologic study in cadaver spines. *Spine* 16(6):629–634. doi:[10.1097/00007632-199106000-00006](https://doi.org/10.1097/00007632-199106000-00006)
22. Tian L (2005) On confidence intervals of a common intraclass correlation coefficient. *Stat Med* 24(21):3311–3318. doi:[10.1002/sim.2145](https://doi.org/10.1002/sim.2145)
23. Vernon-Roberts B, Pirie CJ (1977) Degenerative changes in the intervertebral discs of the lumbar spine and their sequelae. *Rheumatol Rehabil* 16(1):13–21