Annette Kettler
Friederike Rohlmann
Cornelia Neidlinger-Wilke
Karin Werner
Lutz Claes
Hans-Joachim Wilke

Validity and interobserver agreement of a new radiographic grading system for intervertebral disc degeneration: Part II. Cervical spine

Received: 21 January 2005
Revised: 23 August 2005
Accepted: 4 September 2005
Published online: 14 April 2006
© Springer-Verlag 2006

Abstract A new radiographic grading system for a more objective assessment of lumbar intervertebral disc degeneration has been described and tested in Part I of this study. The aim of the present Part II of the study was to adapt this system to the cervical spine, and to test it for validity and interobserver agreement. Some modifications of the grading system described in Part I were necessary to make it applicable to the cervical spine. Its basic structure, however, stayed untouched. The three variables “Height Loss”, “Osteophyte Formation” and “Diffuse Sclerosis” first have to be graded individually. Then, the “Overall Degree of Degeneration” is assigned on a four-point scale from 0 (no degeneration) to 3 (severe degeneration). For validation, the radiographic degrees of degeneration of 28 cervical discs were compared to the respective macroscopic ones, which were defined as “real” degrees of degeneration. The interobserver agreement was determined between one experienced and one unexperienced observer using the radiographs of 57 cervical discs. Quadratic weighted Kappa coefficients (κ) with 95% confidence limits (95% CL) were used for statistical evaluation. The validation of the new version of the radiographic

grading system showed a moderate agreement with the “real”, macroscopic overall degree of degeneration ($\kappa = 0.599$, 95% CL 0.421–0.786). In 64% of all discs the “real” overall degree of degeneration was underestimated but never overestimated. This underestimation, however, was much less pronounced and the Kappa coefficients were significantly higher for the three variables: Height Loss, Osteophyte Formation, and Diffuse Sclerosis separately. The agreement between the radiographic ratings of the experienced and the unexperienced observer was substantial for the overall degree of degeneration ($\kappa = 0.688$, 95% CL 0.580–0.796), almost perfect for the variable, Height Loss, moderate for Osteophyte Formation and fair for Diffuse Sclerosis. In conclusion, we believe that the new version of the radiographic grading system is a sufficiently valid and reliable tool to quantify the degree of degeneration of individual cervical intervertebral discs. In comparison to the version for the lumbar spine described in Part I, however, a slightly higher tendency to underestimate the “real” overall degree of degeneration and somewhat higher interobserver differences have to be expected.

Part I of this article can be found at <http://dx.doi.org/10.1007/s00586-005-1029-9>

A. Kettler · C. Neidlinger-Wilke
K. Werner · L. Claes · H.-J. Wilke (✉)
Institute of Orthopaedic Research
and Biomechanics, University of Ulm,
Helmholtzstrasse 14, 89081 Ulm, Germany
E-mail: hans-joachim.wilke@uni-ulm.de

F. Rohlmann
Department of Biometry and Medical
Documentation, University of Ulm,
Ulm, Germany

Introduction

The morphology of cervical disc degeneration has often been described in literature [3, 9, 11, 16, 24]. Based on these descriptive studies, several classification systems have been developed specifically for the cervical spine [2, 6, 10, 13, 15, 19, 21, 25]. Of these, however, only four are based on plain radiographs [2, 10, 21]. This number is very small in view of the fact that radiographic grading systems have several advantages: in contrast to macroscopic and histologic systems, they can also be applied in vivo; in contrast to discographic systems, they are less invasive; and in contrast to grading systems based on magnetic resonance images or computed tomography scans, they only require a standard X-ray machine and are less expensive. But more than their small number, missing reliability and validity tests reduce their applicability. The only exception is the grading system of Kellgren et al. [10], which has been tested for reliability by Cote et al. [4].

Another critical point of these grading systems is their terminology. The systems of Kellgren and Lawrence [11] and Kellgren et al. [10] are based on terms such as “minimal”, “definite”, “moderate”, or “severe”, and the system of Brooker and Barter [2] is using the terms “slight”, “moderate” and “severe” to describe the degree of changes seen on the radiographs. Such terms are not well defined and easily decrease the objectivity of the grading system. This effect was shown in Part I of this study in which a more objective radiographic grading system for disc degeneration was developed. This new grading system, however, was specifically adapted to the intervertebral discs of the lumbar spine

and should therefore not be used to grade those of the cervical spine.

The aim of this present Part II study was to adapt this system to the cervical spine, and to test it for validity and agreement between observers with different levels of experience.

Materials and methods

Some modifications of the grading system described in Part I of this study were necessary to make it applicable to the cervical spine. First of all, for the cervical spine, only lateral films are needed, since postero-anterior films of the cervical spine are difficult to evaluate and do not add to the information provided by lateral films.

Despite this modification, the basic structure of the new grading system stayed the same: the three variables, “Height Loss”, “Osteophyte Formation” and “Diffuse Sclerosis” first have to be graded individually on a scale from 0 to 3, then, based on the sum of these three scores, the overall degree of degeneration is assigned to each disc on a four-point scale from 0 (no degeneration) to 3 (severe degeneration) (Table 1).

No modifications were necessary concerning the variable Height Loss. In both versions, this variable is defined as the average anterior and posterior decrease in disc height referred to the respective height before degeneration. For its assessment, first, the actual anterior and posterior disc height has to be determined. For this purpose, the distance between the four edges of the adjacent vertebral bodies and the midplane of the intervertebral space is determined and referred to the

Table 1 New version of the radiographic grading system for intervertebral disc degeneration, which is adapted to the cervical spine

Radiographic grading system for cervical intervertebral disc degeneration (based on lateral radiographs^a)

Height Loss	Osteophyte Formation ^a	Diffuse Sclerosis ^a	Overall Degree of Degeneration
Anterior and posterior Height Loss with respect to the individual height before degeneration	Sum of points of 4 edges: no osteophytes: 0 points < 2 mm: 1 point ≥ 2 mm but < 4 mm: 2 points ≥ 4 mm: 3 points	Sum of points of both adjacent vertebral bodies: no sclerosis: 0 points 1/2 partially or completely affected: 1 point > 1/2 partially or completely affected: 2 points	Sum of points of Height Loss, Osteophyte Formation and Diffuse Sclerosis
0 = 0%	0 = 0 points	0 = 0 points	0 point = Grade 0 (no degeneration)
1 = < 33%	1 = 1–4 points	1 = 1 point	1–3 points = Grade 1 (mild degeneration)
2 = ≥ 33%	2 = 5–8 points	2 = 2 points	4–6 points = Grade 2 (moderate degeneration)
but < 66%	3 = 9–12 points	3 = 3–4 points	7–9 points = Grade 3 (severe degeneration)
3 = ≥ 66%			

The three variables Height Loss, Osteophyte Formation and Diffuse Sclerosis are first graded individually on a scale from 0 to 3. The overall degree of degeneration is then assigned according to the sum of these three scores

^aDifferent from the original version for the lumbar spine (see Part I)

antero-posterior diameter of the cranial vertebral body (Fig. 1). This procedure does not have to be carried out using drawings or digitisation and, thus, is meant to support a rough estimation of actual anterior and posterior disc height. In a second step, the actual anterior and posterior disc height is then compared to the respective individual height before degeneration, which is estimated based on the normal values reported by Frobin et al. [7]. These very precise values specifically refer to the anterior intervertebral height, which is, if assessed according to Frobin et al. [7] independent from image distortion (Table 2). To account for interindividual differences, the ranges of normal disc height

Fig. 1 To assess the degree of “Height Loss”, first, the actual disc height has to be determined. For this purpose the anterior and posterior edges of the adjacent vertebral bodies (*small white circles*) are defined as those points having the largest distance to the centre of the vertebral body (*black points*). Then, the distance of each of these four edges to the midplane of the disc (*dashed line*) is measured. The sum of the two anterior distances finally is defined as the actual anterior disc height, the sum of the two posterior distances as the actual posterior disc height. This procedure is meant to support the estimation of actual disc height but does not have to be carried out using drawings or digitisation. In a second step, this actual height is compared to the respective height before degeneration, which is estimated based on the normal values reported by Frobin et al. 1997 (Table 2)

Table 2 Normal values of anterior disc height normalised to the antero-posterior diameter of the cranial vertebral body (=100%) (mean of male and female subjects according to Frobin et al. (2002))

Normal values of anterior disc height (modified according to Frobin et al. 2002)

	Mean (%)	Mean -2 SD (%)	Mean +2 SD (%)
C2-3	34	22	45
C3-4	38	26	51
C4-5	39	28	50
C5-6	38	29	47
C6-7	34	23	44

Mean mean value; *SD* standard deviation

should be used for comparison rather than their mean value. In contrast, precise normal values for posterior intervertebral height are not available. They have to be assumed to be smaller or as high but not higher than the normal values for anterior height.

To assess the variable, Osteophyte Formation, the number of osteophytes is counted and their length is measured in both versions of the grading system. However, in the cervical version, only the two anterior and the two posterior but not the four lateral edges of the adjacent vertebral bodies are evaluated. Furthermore, the length needed to assign one point was reduced from 3 to 2 mm and for two points from 6 to 4 mm.

The assessment of Diffuse Sclerosis, also had to be slightly modified. Even though, in both versions the number of sclerotic regions is counted, the adjacent halves of the upper and lower vertebral body are each divided into only two but not into four regions for the cervical spine (Fig. 2). Thickened bony endplates should be counted only if their border is diffuse.

For validation, the radiographic degrees of degeneration of 28 cervical intervertebral discs from eight fresh frozen human cadaveric osteoligamentous spine specimens, all without known history of trauma or spinal disease were compared to the respective macroscopic ones, which were defined as “real” degrees of degeneration (Table 3, steps 2 and 3). The specimens were first X-rayed in lateral direction (43805 X-Ray System, Faxitron Series, Hewlett Packard, USA; film to source distance 61 cm) using a tube voltage of 40–45 kV and an exposure time of 5 min. Then, still being frozen they were cut in the midsagittal plane. The cutting surfaces were photographed and stored for evaluation. In order to allow a direct comparison between the radiographic and the macroscopic ratings, the macroscopic grading systems also covered the three variables: Height Loss, Osteophyte Formation and Diffuse Sclerosis. For the assessment of the variable, Diffuse Sclerosis, however, the adjacent halves of the upper and lower vertebral body were each divided into four regions on the macroscopic slices but not into two as on the radiographs.

assessment of “Diffuse Sclerosis”

Fig. 2 To evaluate the variable “Diffuse Sclerosis” in the assessment of cervical disc degeneration, the lower half of the upper vertebral body and the upper half of the lower vertebral body are each divided into two regions. Then, the number of regions is counted, which are covered by sclerosis. Note that a partially covered region is counted as if it was completely covered

To be as close to the “real” degree of degeneration as possible, another three variables were added to the macroscopic grading system: “Nucleus Pulposus”,

“Annulus Fibrosus” and “Endplate Cartilage” (Table 4; modified according to Thompson et al. [23]).

This macroscopic grading system was first checked for interobserver reliability using 63 discs from 16 donors (Table 3, step 1). Out of these 63 discs, 28, for which radiographs were available, were additionally used to define the “real” degree of degeneration (Table 3, step 2). For these two steps of evaluation, two observers, who were both familiar with disc degeneration and who had several years of experience in spinal research evaluated the photographs of the cutting surfaces independently from each other. The “real” degree of degeneration of the 28 discs was then defined as the mean value of the results of both observers. Then, all these 28 discs were additionally graded radiographically by one of these two observers (Table 3, step 3). In order to ensure that this observer was not biased by the macroscopic grading carried out a few days before, the radiographs were blinded and put in a randomised order. Statistically, the agreement between the radiographic and the “real” macroscopic degrees of degeneration was determined using weighted Kappa coefficients (quadratic weights) with 95% confidence limits (95% CL). They were calculated according to Fleiss and Cohen [5] using the software SAS 8.2 [18]. For this statistical evaluation an independency of the observation of each intervertebral disc was assumed.

In order to show whether the grade assigned to individual discs depends on the degree of experience of the observer, the agreement between one experienced and one unexperienced observer was determined. Both observers graded the lateral radiographs of 17 mono- or polysegmental osteoligamentous cervical spine specimens with an overall of 57 intervertebral discs (Table 3, step 4). The experienced observer was one of the two who were involved in the validation of the new version

Table 3 Overview on the steps of evaluation

Step	Specimens				Grading system	Observer	Aim
	Number of discs	Number of donors	Gender	Mean (min–max) age			
1	63	16	11 female, 5 male	78 (59–92) years	Macroscopic	Two experienced observers	Interobserver agreement (see Table 5)
2	Specimens of step 1, for which radiographs were available 28	8	6 female, 2 male	78 (67–92) years	Macroscopic	Two experienced observers	Definition of “real” degree of degeneration
3	Same specimens as in step 2 28	8	6 female, 2 male	78 (67–92) years	Radiographic	One experienced observer	Validation (comparison to “real” degree of degeneration) (see Table 6, Fig. 3)
4	57	17	11 female, 6 male	77 (61–92) years	Radiographic	One experienced observer One unexperienced observer	Interobserver agreement (see Table 7, Fig. 4)

Table 4 Macroscopic grading system for cervical intervertebral disc degeneration used as “gold standard” to define the “real” degree of degeneration (modified according to the systems found in literature)

Macroscopic grading system for lumbar intervertebral disc degeneration (based on midsagittal sections)			
Nucleus Pulposus	Annulus Fibrosus	Endplate Cartilage	
Appearance	Appearance	Sum of points of both cartilaginous endplates: normal: 0 points thickness irregular: 1 point focal defect(s): 2 points (almost) complete destruction: 3 points	
0 = bulging gel 1 = fibrous tissue; loss of annular-nuclear demarcation 2 = focal clefts 3 = complete disruption of nucleus or complete transformation into tissue other than nucleus tissue	0 = discrete fibrous lamellas 1 = mucinous infiltration 2 = focal disruptions 3 = complete disruption of anterior plus posterior annulus or complete transformation into tissue other than annulus tissue	0 = 0 points 1 = 1–2 points 2 = 3–4 points 3 = 5–6 points	
Height Loss	Osteophyte Formation	Diffuse Sclerosis	Overall Degree of Degeneration
Anterior and posterior Height loss with respect to the individual height before degeneration	Sum of points of 4 edges: no osteophytes: 0 points < 2 mm: 1 point ≥ 2 mm but < 4 mm: 2 points ≥ 4 mm: 3 points	Sum of points of both adjacent vertebral bodies: no sclerosis: 0 points 1/4 partially or completely affected: 1 point 2/4 partially or completely affected: 2 points > 2/4 partially or completely affected: 3 points	Sum of points of Nucleus Pulposus, Annulus Fibrosus, Endplate Cartilage, Height Loss, Osteophyte Formation and Diffuse Sclerosis
0 = 0% 1 = < 33% 2 = ≥ 33 but < 66% 3 = ≥ 66%	0 = 0 points 1 = 1–4 points 2 = 5–8 points 3 = 9–12 points	0 = 0 points 1 = 1–2 points 2 = 3–4 points 3 = 5–6 points	0 point = Grade 0 (no degeneration) 1–6 points = Grade 1 (mild degeneration) 7–12 points = Grade 2 (moderate degeneration) 13–18 points = Grade 3 (severe degeneration)

The six variables Nucleus Pulposus, Annulus Fibrosus, Endplate Cartilage, Height Loss, Osteophyte Formation and Diffuse Sclerosis are first graded individually on a scale from 0 to 3. The overall degree of degeneration is then assigned according to the sum of

these six scores. Note that the macroscopic and the radiographic grading system are identical for the variables Height Loss and Osteophyte Formation but slightly differ for Diffuse Sclerosis

of the grading system. The unexperienced observer was a mechanical engineer, who had never had any medical training. This unexperienced observer, however, was trained before grading the discs. The grading system was explained using some training radiographs and the radiographic appearance of the most common spinal diseases was demonstrated in a 30 min session. Then, written instructions were handed over, in which the assessment of the three variables was again explained and the normal values of anterior disc height were listed.

Besides these instructions the unexperienced observer did not get any further help during grading.

Statistically, the agreement between the ratings of the experienced and the unexperienced observer was evaluated using the same type of weighted Kappa coefficient earlier used for validation [5]. For both, the validation and the assessment of the interobserver agreement, a κ of < 0.00 was interpreted as poor agreement, 0.00–0.20 as slight agreement, 0.21–0.40 as fair agreement, 0.41–0.60 as moderate agreement,

0.61–0.80 as substantial agreement, and >0.81 as almost perfect agreement [14].

Results

The agreement between the macroscopic ratings of the two experienced observers was substantial for Osteophyte Formation ($\kappa=0.631$) and almost perfect for the other five variables (Height Loss, Diffuse Sclerosis, Nucleus Pulposus, Annulus Fibrosus, Endplate Cartilage) and for the overall degree of degeneration (κ between 0.818 and 0.897) (Table 5). Identical overall degrees of degeneration were assigned to 86% of all discs and in the remaining 14%, the differences did not exceed one degree.

The validation of the radiographic grading system revealed an almost perfect agreement with the macroscopic, “real” degree of degeneration for the variable, Height Loss ($\kappa=0.904$) and a slightly lower but still substantial agreement for Osteophyte Formation and Diffuse Sclerosis (κ 0.755, respectively 0.714) (Table 6). For the overall degree of degeneration the agreement was lowest with a Kappa coefficient of 0.599. In this case, the “real” degree of degeneration was underestimated in 64% of all discs (Fig. 3). Since the agreement for the three variables, Height Loss, Osteophyte Formation and Diffuse Sclerosis, was substantial or almost perfect, this underestimation of the overall degree of degeneration has to be attributed to the remaining three variables Nucleus Pulposus, Annulus Fibrosus, Endplate Cartilage which could only be evaluated on the macroscopic slices but not on the radiographs.

The agreement between the radiographic ratings of the experienced and the unexperienced observer was almost perfect for the variable, Height Loss ($\kappa=0.827$), moderate for Osteophyte Formation ($\kappa=0.559$) and fair for Diffuse Sclerosis ($\kappa=0.310$) (Table 7). For the

Table 5 Agreement between the macroscopic ratings of the two experienced observers (weighted Kappa coefficients with 95% confidence limits)

Interobserver agreement macroscopy ($n=63$ cervical intervertebral discs)			
	κ	95% confidence limits	
		Lower	Upper
Height Loss	0.828	0.748	0.908
Osteophyte Formation	0.631	0.481	0.781
Diffuse Sclerosis	0.841	0.736	0.945
Nucleus Pulposus	0.876	0.820	0.931
Annulus Fibrosus	0.852	0.796	0.907
Endplate Cartilage	0.818	0.739	0.898
Overall grade	0.897	0.828	0.966

Table 6 Agreement between the radiographic and the macroscopic, “real” degrees of degeneration of 28 cervical intervertebral discs (weighted Kappa coefficients with 95% confidence limits)

Agreement between radiography and macroscopy ($n=28$ cervical intervertebral discs)			
	κ	95% confidence limits	
		Lower	Upper
Height Loss	0.904	0.823	0.985
Osteophyte Formation	0.755	0.585	0.924
Diffuse Sclerosis	0.714	0.338	1.090
Overall grade ^a	0.599	0.412	0.786

^a Note that the overall degree of degeneration covers only three variables in the radiographic (Height Loss, Osteophyte Formation, Diffuse Sclerosis) but six variables in the macroscopic grading system (additionally Nucleus Pulposus, Annulus Fibrosus, Endplate Cartilage)

overall degree of degeneration the agreement was substantial with a κ of 0.688; however, the unexperienced observer tended to assign lower degrees of degeneration to the discs than the experienced one (Fig. 4). Nevertheless, most ratings of the two observers were identical: The same degree of Height Loss was assigned to 63% of all discs, the same degree of Osteophyte Formation to 54%, the same degree of Diffuse Sclerosis to 70% and the same overall degree of degeneration to 61%. The differences in grade assignment were never higher than one degree except for the variable Diffuse Sclerosis. In this case, eight discs differed by two degrees. Differences of more than two degrees were not observed.

Discussion

In Part I of this study, the radiographic grading systems for lumbar intervertebral disc degeneration available from literature were combined to a new one, in which terms such as mild, severe, small or large were replaced by better defined and more objective ones. In the present Part II, this system was modified to be applicable to the cervical spine. Then, the new version was tested for validity and interobserver agreement in the same way as to the lumbar version in Part I.

In some aspects the interobserver agreement of the new radiographic grading system differed between the present version for the cervical and the first version for the lumbar spine. Except for the variable Height Loss the interobserver agreement was lower for the cervical than for the lumbar spine. This difference was most pronounced for the variable, Diffuse Sclerosis, where κ was only 0.310 for the cervical but 0.681 for the lumbar version. Only this fair agreement for the cervical version can be explained, first, by the small dimensions of the

Fig. 3 Agreement between the radiographic and the macroscopic, “real” degree of degeneration of 28 cervical intervertebral discs. Each field contains the number of discs rated with 0, 1, 2, or 3 points radiographically (rating of one experienced observer) and with 0, 0.5, 1, 1.5, 2, 2.5, or 3 points macroscopically (mean value of the ratings of two experienced observers)

Table 7 Agreement between the radiographic ratings of one experienced and one unexperienced observer (weighted Kappa coefficients with 95% confidence limits)

Interobserver agreement radiography (<i>n</i> = 57 cervical intervertebral discs)			
	κ	95% confidence limits	
		Lower	Upper
Height Loss	0.827	0.747	0.907
Osteophyte Formation	0.559	0.404	0.714
Diffuse Sclerosis	0.310	-0.030	0.650
Overall grade	0.688	0.580	0.796

cervical vertebrae since sclerosis can hardly be assessed in small areas. Furthermore, on lateral radiographs of the cervical spine, the posterior halves of the vertebral bodies are almost completely covered by the lateral processes. Thus, in the cervical spine, sclerosis additionally has to be distinguished from a decrease in X-ray transparency caused by overlying structures. In contrast, in the lumbar spine, the lateral processes are much smaller and, thus, cover much less of the vertebral body.

Besides the variable, Diffuse Sclerosis, the assessment of osteophytes also was characterised by a lower interobserver agreement for the cervical than for the lumbar spine. This may be caused, first, by the special shape of the cervical vertebrae. On lateral films, the anterior, inferior corner of a cervical vertebra often has the shape of a “nose” pointing in antero-inferior direction. Whether this nose still is normal or not often is difficult to decide. Second, similarly to the difficulties concerning the assessment of sclerosis, the lateral processes complicate the assessment of osteophytes since they generally cover the two posterior edges of the vertebral bodies. These differences in interobserver agreement between the two versions of the grading system also indicate that the quality of the radiographs may play an important role. For this reason, postero-anterior radiographs were excluded from the assessment of cervical disc degeneration: such films can hardly be evaluated since most discs are generally not well aligned. Besides the physiologic lordotic alignment of the cervical spine, scoliotic or other deformities may even complicate the evaluation of the lateral but also of the postero-anterior films. In cases of deformity the grading system should therefore be applied only very cautiously. Since the quality of the radiographs also depends on whether patients or specimens are X-rayed the agreement measured in this study may slightly differ from that expected in clinical practice. Slight differences are also expected since only one experienced and one unexperienced observer graded the discs and since the number of discs examined was restricted. Thus,

HEIGHT LOSS

unexperienced obs.

experienced obs.

pts.:	0	1	2	3	total
0	19	2	0	0	21
1	8	7	4	0	19
2	0	4	5	1	10
3	0	0	2	5	7
total	27	13	11	6	57

OSTEOPHYTE FORMATION

unexperienced obs.

experienced obs.

pts.:	0	1	2	3	total
0	14	12	0	0	26
1	4	12	1	0	17
2	0	9	5	0	14
3	0	0	0	0	0
total	18	33	6	0	57

DIFFUSE SCLEROSIS

unexperienced obs.

experienced obs.

pts.:	0	1	2	3	total
0	39	5	2	0	46
1	1	1	0	0	2
2	6	1	0	0	7
3	0	0	2	0	2
total	46	7	4	0	57

OVERALL GRADE

unexperienced obs.

experienced obs.

pts.:	0	1	2	3	total
0	9	6	0	0	15
1	6	21	1	0	28
2	0	4	5	0	9
3	0	0	5	0	5
total	15	31	11	0	57

Fig. 4 Agreement between the radiographic ratings of one experienced and one unexperienced observer. Each field contains the number of cervical intervertebral discs rated with the respective scores

finally, the agreement measured in this study has to be interpreted as being a trend, that most probably slightly varies depending on the circumstances under which the grading system is used.

In contrast to the lumbar spine, the interobserver agreement of the version for the cervical spine widely differed between the three variables: Height Loss, Osteophyte Formation and Diffuse Sclerosis. Therefore it is recommended to report each of these variables separately instead of the overall degree of degeneration only. This approach is also advantageous in view of the small number of only three variables. A disc with a severe Height Loss but only small osteophytes, for example, may have the same overall degree of degeneration as a disc with only little Height Loss but large osteophytes. Thus, in this example, the overall degree of degeneration would indicate the same severity, whereas the individual appearances of the discs were very different.

Nevertheless, for the new version of the grading system, the agreement between the overall ratings of both observers was still substantial ($\kappa = 0.688$). Unfortunately, there is only little data available for direct comparison since the only radiographic grading system for cervical disc degeneration tested for interobserver agreement is that of Kellgren et al. [10]. The results of these tests, which were carried out by Cote et al. [4], revealed an almost perfect agreement for the overall degree of degeneration ($\kappa = 0.71$). Compared to our new version for the cervical spine, this value is only little higher, even though it reflects the agreement between three medical specialists, who all had 10 or more years of

experience. In view of this fact, the interobserver agreement of the new cervical grading system, which was measured for two observers with large differences in experience, is deemed to be good.

The validation of the new version of the radiographic grading system was carried out in the same way as the validation of the lumbar version in Part I of this study, i.e. the macroscopic degree of degeneration was defined as being “real” since macroscopic slices directly reflect the changes within the disc, whereas radiographs only depict the surrounding bony structures. As to the three variables: Height Loss, Osteophyte Formation and Diffuse Sclerosis, the agreement between the radiographic and the macroscopic, “real” degree of degeneration was higher for the cervical than for the lumbar spine with Kappa coefficients of 0.904, 0.755 and 0.714. Thus, the new cervical version of the radiographic grading system is deemed to be almost valid for each of these single variables (Fig. 5). The use of a less complex and, thus, a less accurate technique in the assessment of actual disc height and the comparison of the actual disc height to normal values obtained in a different technique therefore did not cause any significant error. Concerning the overall degree of degeneration, however, the validity was somewhat lower for the cervical than for the lumbar spine with a Kappa coefficient of 0.599, and the “real” overall degree of degeneration tended to be more often underestimated. If MR images were used instead of radiographs, early changes within the nucleus might become visible before first signs of degeneration become detectable on radiographs. In the past, several grading systems for disc degeneration based on MR images have been proposed; however, none of those for the cervical spine was tested for reliability or validity [8, 15, 17, 19, 20, 22, 25]. Thus, whether MR imaging is really superior

Fig. 5 Examples of the four degrees of degeneration

Grade 0 (no degeneration) (C2-3, male 59 years):

lateral radiograph

midsagittal frozen cut

Grade 1 (mild degeneration) (C5-6, female, 72 years):

lateral radiograph

midsagittal frozen cut

Grade 2 (moderate degeneration) (C4-5, female, 81 years):

lateral radiograph

midsagittal frozen cut

Grade 3 (severe degeneration) (C3-4, female, 92 years):

lateral radiograph

midsagittal frozen cut

to plain radiography is not yet well understood. However, for the lumbar spine Benneker et al. [1] could show that conventional radiography remains a cost-effective, non-invasive in vivo grading method to detect early disc degeneration and, combined with MRI, correlates best with morphological and biochemical assessment of disc degeneration.

Conclusions

In conclusion, we believe that the new version of the radiographic grading system is a sufficiently valid and reliable tool to quantify the degree of degeneration of individual cervical intervertebral discs. In comparison to the version for the lumbar spine described in Part I of this study, however, a slightly higher tendency to

underestimate the “real” overall degree of degeneration and somewhat higher interobserver differences have to be expected.

Similar to Part I of this study, focus was put on the agreement between one experienced and one unexperienced observer to evaluate the objectivity of the new version of the grading system. Other parameters such as the intraobserver agreement, the agreement between observers with similar degrees of experience, the agreement between more than only two observers, the agreement between whole institutions, the effect of the quality of the radiographs or the suitability of the grading system for longitudinal studies need to be the aim of future investigations.

Acknowledgement The authors gratefully acknowledge the *Deutsche Arthrose-Hilfe e.V.* for financial support.

References

1. 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(1):27–35
2. Brooker AE, Barter RW (1965) Cervical spondylosis. a clinical study with comparative radiology. *Brain* 88(5):925–936
3. Collins DH (1949) The pathology of articular and spinal diseases. Edward Arnold & Co, London
4. Cote P, Cassidy JD, Yong-Hing K, Sibley J, Loewy J (1997) Apophysial joint degeneration, disc degeneration, and sagittal curve of the cervical spine. Can they be measured reliably on radiographs? *Spine* 22(8):859–864
5. Fleiss J, Cohen J (1973) The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educ Psychol Meas* 33:613–619
6. Friedenberg ZB, Miller WT (1963) Degenerative disc disease of the cervical spine. *J Bone Joint Surg Am* 45:1171–1178
7. Frobin W, Leivseth G, Biggemann M, Brinckmann P (2002) Vertebral height, disc height, posteroanterior displacement and dens-atlas gap in the cervical spine: precision measurement protocol and normal data. *Clin Biomech (Bristol, Avon)* 17(6):423–431
8. Fujiwara A, Lim TH, An HS, 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
9. Hirsch C (1972) Some morphological changes in the cervical spine during ageing. In: Hirsch C, Zotterman Y (eds) *Cervical pain*. Pergamon, New York, pp 21–32
10. Kellgren JH, Jeffrey MR, Ball J (1963) The epidemiology of chronic rheumatism. Vol. II: Atlas of standard radiographs of arthritis. Blackwell, Oxford, pp 14–19
11. Kellgren JH, Lawrence JS (1952) Rheumatism in miners. II. X-ray study. *Br J Ind Med* 9(3):197–207
12. Kellgren JH, Lawrence JS (1957) Radiological assessment of osteoarthritis. *Ann Rheum Dis* 16(4):494–502
13. Kettler A and Wilke H-J (2005) Review of existing grading systems for cervical and lumbar disc and facet joint degeneration. *Eur Spine J* (accepted)
14. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33(1):159–174
15. Lehto IJ, Terti MO, Komu ME, Paajanen HE, Tuominen J, Kormano MJ (1994) Age-related MRI changes at 0.1 T in cervical discs in asymptomatic subjects. *Neuroradiology* 36(1):49–53
16. Macnab I (1975) Cervical spondylosis. *Clin Orthop* 109:69–77
17. Pfirrmann CW, Metzger A, Zanetti M, Hodler J, Boos N (2001) Magnetic resonance classification of lumbar intervertebral disc degeneration. *Spine* 26(17):1873–8
18. SAS (1999) SAS Institute Inc.: Cary, NC, USA
19. Schellhas KP, Smith MD, Gundry CR, Pollei SR (1996) Cervical discogenic pain. Prospective correlation of magnetic resonance imaging and discography in asymptomatic subjects and pain sufferers. *Spine* 21(3):300–311; discussion 311–312
20. Schneiderman G, Flannigan B, Kingston S, Thomas J, Dillin WH, Watkins RG (1987) Magnetic resonance imaging in the diagnosis of disc degeneration: correlation with discography. *Spine* 12(3):276–281
21. Silberstein CE (1965) The evolution of degenerative changes in the cervical spine and an investigation into the “Joints of Luschka”. *Clin Orthop* 40:184–204
22. Terti M, Paajanen H, Laato M, Aho H, Komu M, Kormano M (1991) Disc degeneration in magnetic resonance imaging. A comparative biochemical, histologic, and radiologic study in cadaver spines. *Spine* 16(6):629–634
23. Thompson JP, Pearce RH, Schechter MT, Adams ME, Tsang IK, Bishop PB (1990) Preliminary evaluation of a scheme for grading the gross morphology of the human intervertebral disc. *Spine* 15(5):411–415
24. Töndury G (1972) The behaviour of the cervical discs during life. In: Hirsch C, Zotterman Y (eds) *Cervical pain*. Pergamon, New York, pp 59–66
25. Viikari-Juntura E, Raininko R, Videman T, Porkka L (1989) Evaluation of cervical disc degeneration with ultralow field MRI and discography. An experimental study on cadavers. *Spine* 14(6):616–619