ORIGINAL ARTICLE

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

Christian Liebsch[1] · Youping Tao[1] · Annette Kienle[2] · Hans-Joachim Wilke[1]

Abstract

Purpose The aim of this study was to assess the validity and objectivity of a new quantitative radiographic grading system for thoracic intervertebral disc degeneration.

Methods The new grading system involves the measurement variables "Height loss" and "Osteophyte formation", which are determined from lateral radiographs, resulting in the "Overall degree of degeneration" on a four-point scale from 0 (no degeneration) to 3 (severe degeneration). Validation was performed by comparing the radiographic degrees of degeneration of 54 human intervertebral discs to the respective macroscopic degrees, which were defined as the "real" degrees of degeneration. Interobserver agreement was examined using radiographs of 135 human thoracic intervertebral discs. Agreement was quantified by means of quadratically weighted Kappa coefficients with 95% confidence limits (CL).

Results Validation revealed almost perfect agreement between the radiographic and the macroscopic overall degrees of degeneration (Kappa 0.968, CL 0.944–0.991), while the macroscopic grades tended to be underestimated in low degeneration grades. Radiographic grading of two independent observers also exhibited almost perfect agreement (Kappa 0.883, CL 0.824–0.941) as well as tendencies towards rater-dependent differences in low degeneration grades.

Conclusion The new quantitative radiographic grading scheme represents a valid, reliable, and almost objective method for assessing the degree of degeneration of individual thoracic intervertebral discs. Potential effects of interindividual variations and the radiographic superimposition of anatomical structures represent a limitation of this method should be taken into account when using the grading system for clinical and experimental purposes, especially with regard to specific morphological as well as patient- and donor-specific characteristics.

Keywords Thoracic spine · Intervertebral disc · Degeneration grading system · Interobserver reliability · Validity

Introduction

Intervertebral disc degeneration represents a principal aspect when assessing pathological, morphological, and biomechanical changes of the spine. While degenerative changes of the intervertebral disc occur in the entire spine [1], existing grading systems for disc degeneration have been mainly developed for the cervical and lumbar spine [2]. However, thoracic intervertebral discs are more flat and exhibit a different morphology compared to discs of the cervical or lumbar spine [3], potentially leading to differences in modality and velocity of the degeneration process and therefore necessitating a specific grading system.

Previously reported evaluation techniques for thoracic intervertebral disc degeneration were based on pure qualitative characteristics regarding disc height and osteophyte formation [4, 5], while changes need to be quantified in order to ensure repeatability and reproducibility of the measurements. Moreover, grading systems for intervertebral disc degeneration should be validated and tested for objectivity in order to assure clinical applicability and data comparability. Validation of a grading system for thoracic disc degeneration, however, was solely reported in the histological classification of Rutges et al. [6], which is not applicable in clinical practice. Interobserver reliability was determined in the studies of Douvier et al. [7], using the radiographic grading system for the lumbar spine developed by Lane et al. [8], and

✉ Hans-Joachim Wilke
hans-joachim.wilke@uni-ulm.de

1   Institute of Orthopaedic Research and Biomechanics,
     Trauma Research Centre Ulm, Ulm University, Ulm,
     Germany

2   SpineServ GmbH & Co. KG, Ulm, Germany

Raininko et al. [9], performing magnetic resonance imaging for grading. While magnetic resonance imaging reveals several advantages compared to radiography, such as the visualization of cartilaginous tissue and the radiation-free methodology, plain radiographs still provide higher clinical practicability due to inexpensiveness, low time consumption, and high availability in case of follow-up treatments, thus being most appropriate for clinical evaluations and experimental studies.

Quantitative, radiography-based grading schemes for intervertebral disc degeneration have already been developed, validated, and reliability tested for the lumbar spine by Wilke et al. [10] and for the cervical spine by Kettler et al. [11], representing Part I and Part II of an overall classification of intervertebral disc degeneration. In order to complement this classification by means of Part III for the thoracic spine, the purpose of this study was to adapt these two established grading systems to the thoracic spine and to determine the validity and objectivity of the new grading system.

Materials and methods

The new grading system for thoracic intervertebral disc degeneration (Table 1) was based on the grading scheme for the lumbar spine introduced by Wilke et al. [10], assigning an "Overall degree of degeneration" to the disc on a four-point scale from 0 (no degeneration) to 3 (severe degeneration), while some modifications were performed in order to comply with the specific requirements for evaluations of thoracic intervertebral discs.

First of all, solely lateral radiographic images are evaluated, since antero-posterior images of the thoracic spine are difficult to analyse in case of severe disc degeneration, the

presence of the costovertebral joints, and the superimposition by the anterior rib cage structures. This approach conforms to the grading scheme for cervical disc degeneration published by Kettler et al. [11].

Another modification involves the variable "Height loss", which is now evaluated based on the normal values for anterior and posterior disc heights reported by Kunkel et al. [12] and adjacent endplate lengths reported by Panjabi et al. [13] (Fig. 1). In contrast to the grading schemes for cervical and lumbar discs [10, 11], no midplane line is used to be in accordance with the evaluation method of Kunkel et al. [12]. In order to give more weight to this variable, four parameters are evaluated in total for the levels T1-T2 to T11-T12 and two parameters for the level C7-T1 (Table 2).

The determination of the variable "Osteophyte formation" was reduced to length measurement of the two potential anterior osteophytes (Fig. 1), since posterior osteophytes were found to be generally small and thus hardly evaluable in the thoracic spine and furthermore superimposed by the costovertebral joints in radiographic imaging. Moreover, the osteophyte lengths defining the single degrees of osteophyte formation were set to 2.5 and 5 mm, respectively, corresponding to the means of the osteophyte lengths in the grading systems of cervical (2 and 4 mm) [11] and lumbar (3 and 6 mm) [10] intervertebral disc degeneration (Table 1).

Lastly, the variable "Diffuse sclerosis" was not included in the grading system for thoracic intervertebral disc degeneration, since the respective area of analysis can be superimposed by ribs in radiographic imaging, potentially affecting the evaluation quality.

For the validation of the new grading system, the radiographic degrees of degeneration of 54 human thoracic intervertebral discs from 28 donors with a donor age ranging between 37 and 86 years (mean 59 years) were compared

Table 1 New radiographic grading system for thoracic intervertebral disc degeneration (based on lateral radiographs) modified according to the systems found in the literature

Height loss	Osteophyte formation	Overall degree of degeneration
Anterior and posterior height loss with respect to the average height before degeneration	Sum of points of the two anterior edges No osteophytes: 0 points <2.5 mm: 1 point ≥2.5 mm but <5 mm: 2 points ≥5 mm: 3 points	Sum of points of "Height loss" and "Osteophyte formation"
0 = 0%	0 = 0 points	0 point = grade 0 (no degeneration)
1 = <33%	1 = 1–2 points	1–2 points = grade 1 (mild degeneration)
2 = ≥33 but <66%	2 = 3–4 points	3–4 points = grade 2 (moderate degeneration)
3 = ≥66%	3 = 5–6 points	5–6 points = grade 3 (severe degeneration)

The two variables "Height loss" and "Osteophyte formation" are first graded individually on a scale from 0 to 3. The "Overall degree of degeneration" is assigned according to the sum of these two scores

Assessment of actual disc height

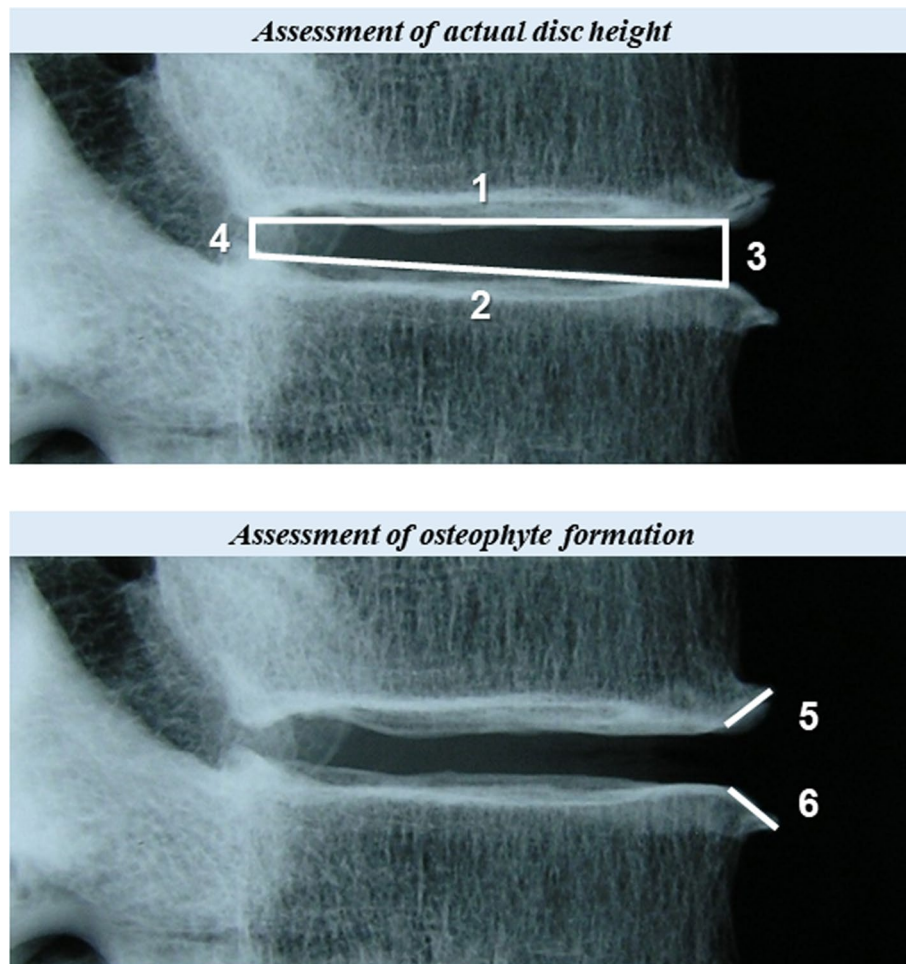**Assessment of osteophyte formation**

Fig. 1 To assess the degrees of height loss and osteophyte formation, six parameters are determined in total. First, pixel values are converted into mm values using the radiopaque scale on the radiograph. The actual disc height is then measured with regard to the antero-posterior disc length. For this purpose, the antero-posterior diameter of the caudal endplate of the cranial vertebral body (1), the antero-posterior diameter of the cranial endplate of the caudal vertebral body (2), as well as the anterior (3) and posterior (4) disc heights are determined. In the second step, the actual disc height is calculated as the four quotients of the two disc heights and the two diameters (3/1, 3/2, 4/1, 4/2) and compared to the respective heights before degeneration, which are estimated based on the combined normal values reported by Kunkel et al. [12] and Panjabi et al. [13] (Table 2). In case of C7-T1, solely the parameters 2, 3, and 4 are used for calculation. The single differences are then averaged and evaluated according to the grading system shown in Table 1 (column "Height loss"). Finally, the lengths of the cranial (5) and caudal (6) anterior osteophytes (if present) are measured and evaluated according to the grading system presented in Table 1 (column "Osteophyte formation"). Example (T7-T8, female, 53 years): $1 = 30.50$ mm, $2 = 30.92$ mm, $3 = 4.41$ mm, $4 = 2.17$ mm, $5 = 2.49$ mm, $6 = 1.77$ mm. Height loss $= (((14\% - ((4.41$ mm $/ 30.50$ mm$) * 100\%)) / 14\%) + ((14\% - ((4.41$ mm $/ 30.92$ mm$) * 100\%)) / 14\%) + ((9\% - ((2.17$ mm $/ 30.50$ mm$) * 100\%)) / 9\%) + ((9\% - ((2.17$ mm $/ 30.92$ mm$) * 100\%)) / 9\%)) / 4 * 100\% = 9\% \rightarrow$ Height loss $= 1$, Osteophyte formation $= 1$ point $+ 1$ point $= 2$ points $\rightarrow$ Osteophyte formation $= 1$, Overall degree of degeneration $=$ Height loss $+$ Osteophyte formation $= 1$ point $+ 1$ point $= 2$ points $=$ Grade $1 =$ Mild degeneration

to the respective macroscopic ones, which were defined as the "real" degrees of degeneration. For this purpose, fresh-frozen thoracic spinal motion segments including the costovertebral joints and ribs without surrounding soft tissue were first X-rayed in lateral direction using a tube voltage of 46.5 kV, an exposure time of 60 s, and a source-to-film distance of 60 cm (Faxitron 43805 N, Hewlett Packard, Palo Alto, the USA). In the next step, the specimens were cut in the mid-sagittal plane using a diamond band saw (EXAKT Advanced Technologies GmbH, Norderstedt, Germany) in

order to photograph the cutting surfaces and thus create the macroscopic images. For direct comparisons of the radiographic and macroscopic degrees of degeneration, both grading systems covered the variables "Height loss" and "Osteophyte formation" (Tables 1 and 3), while the macroscopic grading system additionally included the variables "Nucleus pulposus", "Annulus fibrosus", and "Endplate cartilage" (Table 3) to characterize the "real" degree of degeneration as specifically as possible (modified according to Thompson et al. [14]). To increase the objectivity of the rating, this

Table 2 Mean values of anterior and posterior disc heights normalised to the antero-posterior diameters of the adjacent vertebral bodies (= 100%) (modified according to Kunkel et al. [12] and Panjabi et al. [13]) serving as normal values for the determination of disc height loss

	ADH/CaudEL	ADH/CranEL	PDH/CaudEL	PDH/CranEL
C7-T1	–	21	–	17
T1-T2	19	19	15	15
T2-T3	15	14	9	9
T3-T4	13	13	9	9
T4-T5	11	12	8	8
T5-T6	13	13	8	8
T6-T7	14	13	8	8
T7-T8	14	14	9	9
T8-T9	15	15	8	8
T9-T10	16	16	10	10
T10-T11	19	18	12	11
T11-T12	18	18	10	10

ADH anterior disc height, *PDH* posterior disc height, *CaudEL* caudal endplate length of cranial vertebral body, *CranEL* cranial endplate length of caudal vertebral body

"real" degree of degeneration was defined as the mean value

of the independent ratings of two observers: observer 1 was a biomechanical engineer, who was practiced in using the radiographic grading system for the lumbar spine [10] and involved in the development of the novel grading system, while observer 2 was a spine surgeon, who was exercised in using the radiographic grading system for the cervical spine [11] and not involved in the development. One of the two observers additionally graded the 54 discs using the radiographic grading system, while the radiographs were blinded and assessed in randomized order to prevent bias regarding the previously evaluated macroscopic images.

For testing the objectivity of the new grading system, both observers already performing validity testing independently graded the lateral radiographs of 135 human thoracic intervertebral discs from 36 donors with a donor age ranging 37 and 86 years (mean 58 years), including the 54 discs from validity testing. Radiographs of the fresh-frozen thoracic spinal motion segments including the costovertebral joints and ribs without surrounding soft tissue were either generated using a Faxitron 43805 N X-ray device (Hewlett Packard, Palo Alto, the USA) with a tube voltage of 46.5 kV, an exposure time of 60 s, and a source-to-film distance of 60 cm, or using an AJEX 140 H portable X-ray source (Ajex Meditech, Gyeonggi-do, Republic of Korea) and a FCR Prima CR 391 RU X-ray film developer (Fujifilm Holdings,

Table 3 Macroscopic grading system for thoracic intervertebral disc degeneration used as the gold standard to define the "real" degree of degeneration (modified according to the systems found in the literature)

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	0 = Discrete fibrous lamellas	0 = 0 points
1 = Fibrous tissue; loss of annular-nuclear demarcation	1 = Mucinous infiltration	1 = 1–2 points
2 = Focal clefts	2 = Focal disruptions	2 = 3–4 points
3 = Complete disruption of nucleus or complete transformation into tissue other than nucleus tissue	3 = Complete disruption of anterior plus posterior annulus or complete transformation into tissue other than annulus tissue	3 = 5–6 points
Height loss	Osteophyte formation	Overall degree of degeneration
Anterior and posterior height loss with respect to the average height before degeneration	Sum of points of the two anterior edges No osteophytes: 0 points < 2.5 mm: 1 point ≥ 2.5 mm but < 5 mm: 2 points ≥ 5 mm: 3 points	Sum of points of "Nucleus pulposus", "Annulus fibrosus", Endplate cartilage", "Height loss" and "Osteophyte formation"
0 = 0%	0 = 0 points	0 point = grade 0 (no degeneration)
1 = < 33%	1 = 1–2 points	1–5 points = grade 1 (mild degeneration)
2 = ≥ 33 but < 66%	2 = 3–4 points	6–10 points = grade 2 (moderate degeneration)
3 = ≥ 66%	3 = 5–6 points	11–15 points = grade 3 (severe degeneration)

The five variables "Nucleus pulposus", "Annulus fibrosus", "Endplate cartilage", "Height loss" and "Osteophyte formation'' are first graded individually on a scale from 0 to 3. The "Overall degree of degeneration" is assigned according to the sum of these five scores

Tokyo, Japan) with a tube voltage of 56 kV, an exposure time of 0.8 s, and a source-to-film distance of 1 m.

The agreements between the radiographic and the macroscopic degrees of degeneration as well as between the radiographic rating of the two observers were statistically determined by calculating quadratically weighted Kappa coefficients and 95% confidence limits according to Fleiss and Cohen [15] using the software SPSS 27 (IBM Corp., Armonk, the USA), assuming independency of the observation of each intervertebral disc. In accordance with the recommendation by Landis and Koch [16], a Kappa coefficient 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 > 0.80 as almost perfect agreement for both the validation and the interobserver reliability assessment.

Results

The independent macroscopic ratings of the two observers revealed an almost perfect agreement for the overall grade of degeneration (Kappa 0.840) as well as for the variables "Height loss" (Kappa 0.805) and "Osteophyte formation" (Kappa 0.820) (Table 4). The interobserver agreement for the variables "Nucleus pulposus", "Annulus fibrosus", and "Endplate cartilage" was slightly lower, however, still substantial (Kappa between 0.698 and 0.799), while solely the lower 95% confidence limit of the variable "Endplate cartilage" corresponded to moderate agreement (Kappa 0.578) (Table 4).

Validation of the radiographic grading system with the macroscopic degree of degeneration exhibited an almost perfect agreement for the overall degree of degeneration (Kappa 0.968) as well as for the other two evaluated variables "Height loss" (Kappa 0.919) and "Osteophyte formation" (Kappa 0.906), while all investigated parameters showed almost perfect agreement within the 95% confidence limits (Table 5). Of the 54 evaluated discs, the "real" overall degree

Table 4 Interobserver agreement between the macroscopic ratings of two observers using $n = 76$ thoracic intervertebral discs (quadratically weighted Kappa coefficients with 95% confidence limits)

	Kappa	95% confidence limits	
		Lower	Upper
Height loss	0.805	0.699	0.910
Osteophyte formation	0.820	0.698	0.942
Nucleus pulposus	0.788	0.648	0.928
Annulus fibrosus	0.799	0.670	0.927
Endplate cartilage	0.698	0.578	0.818
Overall grade	0.840	0.744	0.935

Table 5 Agreement between the radiographic ratings of one observer and the macroscopic ratings of two observers using $n = 54$ thoracic intervertebral discs (quadratically weighted Kappa coefficients with 95% confidence limits)

	Kappa	95% confidence limits	
		Lower	Upper
Height loss	0.919	0.857	0.981
Osteophyte formation	0.906	0.829	0.982
Overall grade	0.968	0.944	0.991

of degeneration was over- and underestimated in five cases, respectively, while underestimation tended to be higher in low degeneration grades (Fig. 2). For both variables "Height loss" and "Osteophyte formation", the number of over- and underestimations were about equal, while these were also primarily detected in low degeneration grades (Fig. 2).

The agreement between the radiographic ratings of the two observers was almost perfect for the overall degree of degeneration (Kappa 0.883) as well as for the two variables "Height loss" (Kappa 0.819) and "Osteophyte formation" (Kappa 0.869), while solely the lower 95% confidence limit of the variable "Height loss" revealed substantial agreement (Table 6). Observer 1 generally tended to assign higher grades when observer 2 gave grades 0 or 1, whereas observer 2 tended to assign grade 2 when observer 1 gave grade 1 (Fig. 3). Nevertheless, the ratings were predominantly concordant for all grades, while no differences higher than one grade were found. The overall degrees of degeneration were identical in 87% of all ratings. For the variables "Height loss" and "Osteophyte formation", the concordance rates were 79% and 83%, respectively.

Discussion

A new quantitative radiographic grading system for thoracic intervertebral disc degeneration based on disc height loss and osteophyte formation is presented. This grading system complements the established radiographic grading systems for the cervical [11] and lumbar [10] spine, exhibiting overall high validity and interobserver reliability. Comparisons of the macroscopic ratings revealed highest agreement values of the variables "Height loss" and "Osteophyte formation", indicating general advantage of quantitative parameters compared with qualitative measures. Within the validation measurements, very high agreement values were identified between the radiographic and the macroscopic "real" degrees of degeneration, while a slight tendency towards underestimation of the variables height loss and osteophyte formation was detected at low level of disc degeneration, which might be explained by
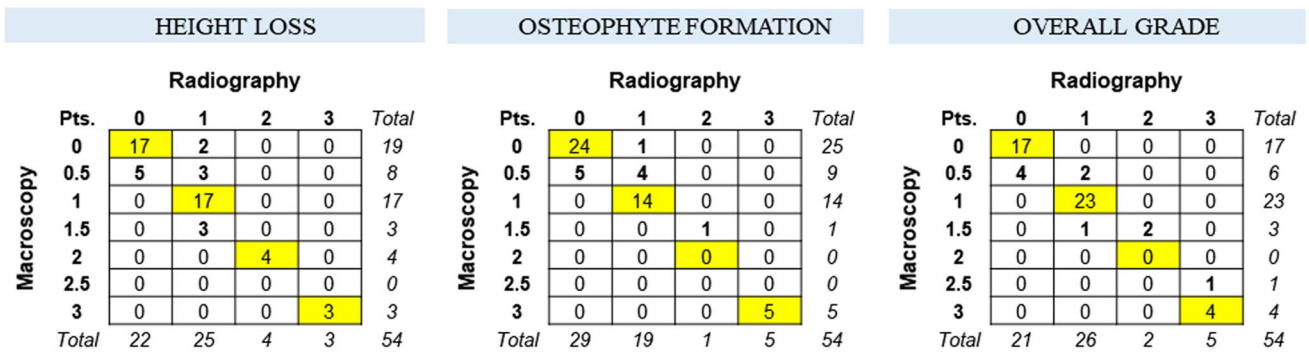
Fig. 2 Agreement between the radiographic and the macroscopic, "real" degree of degeneration using $n=54$ thoracic intervertebral discs. Each field contains the number of discs rated with 0, 1, 2 or 3 points radiographically (rating of one experienced observer) and with 0, 0.5, 1, 1.5, 2, 2.5 or 3 points macroscopically (mean value of the ratings of two observers)

Table 6 Interobserver agreement between the radiographic ratings of two observers using $n=135$ thoracic intervertebral discs (quadratically weighted Kappa coefficients with 95% confidence limits)

	Kappa	95% confidence limits	
		Lower	Upper
Height loss	0.819	0.748	0.890
Osteophyte formation	0.869	0.811	0.928
Overall grade	0.883	0.824	0.941

limited determination specificity regarding small cartilaginous changes when using plane radiographs. Besides, this finding might also be caused by the lower number of moderately and severely compared to non- and mildly degenerated discs evaluated in this study. Beyond that, qualitative comparisons between the radiographic and macroscopic images of the four individual grades support the validity of the new grading system (Fig. 4). Reliability measurements of the radiographic ratings showed very high agreement, indicating an overall high degree of objectivity of the new grading system. However, comparisons between the two raters also revealed minor systematic measurement deviations depending on the degree of degeneration, indicating

individual evaluation preferences that could potentially affect the outcome variability.

Interobserver agreement between the radiographic ratings was generally higher in this study compared to the values of previous studies. Using magnetic resonance imaging, Raininko et al. [9] reported fair agreement for disc height measurement in the mid- and lower thoracic region as well as for osteophyte determination in the mid-thoracic spine and moderate agreement for osteophyte measurements in the lower thoracic region. This might indicate that measurements on plain radiographs can be performed more accurately than on magnetic resonance images. Douvier et al. [7] detected moderate to substantial agreement using plain radiographs using the Lane score [8], which was originally developed for the lumbar spine. Therefore, it can be assumed that the new, radiographic grading system represents an advance in the evaluation of thoracic intervertebral disc degeneration.

The new grading system differs in several aspects from the previously established grading schemes for the cervical [11] and lumbar [10] spine, which were considered as gold standard and modified in order to comply with the specific requirements of thoracic intervertebral discs. The main modifications include the higher weighting of height loss measurements and the non-inclusion of vertebral body
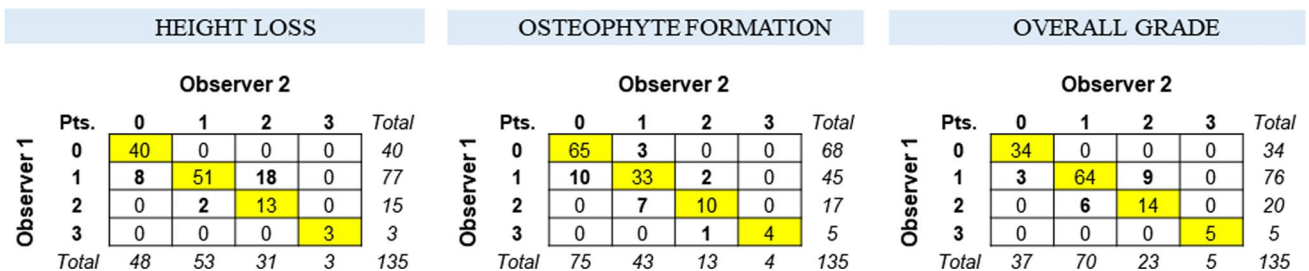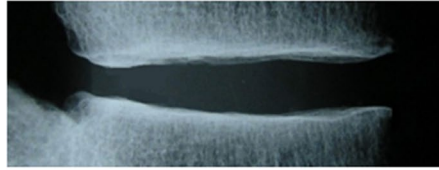

Fig. 3 Interobserver agreement between the radiographic ratings of two observers using $n=135$ thoracic intervertebral discs. Each field contains the number of thoracic intervertebral discs rated with the respective scores
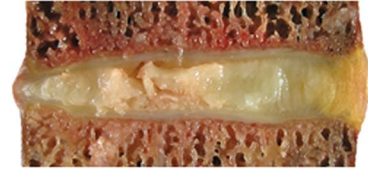
Fig. 4 Examples of the four degrees of thoracic intervertebral disc degeneration

Grade 0 (no degeneration) (T11-T2, female, 49 years):
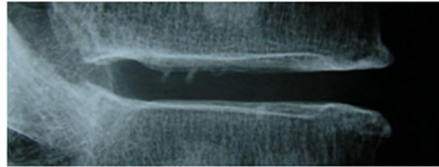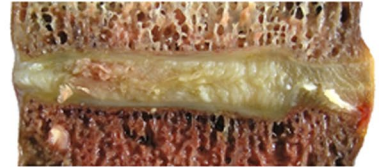*Lateral radiograph* *Midsagittal cut*

Grade 1 (mild degeneration) (T8-T9, male, 54 years):
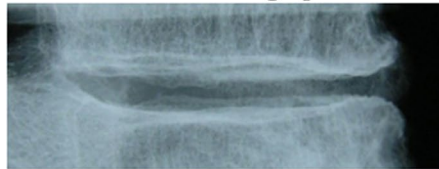*Lateral radiograph* *Midsagittal cut*

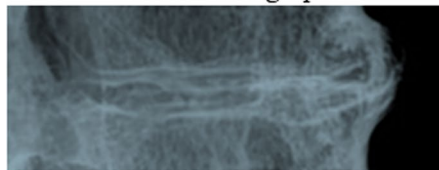**Grade 2 (moderate degeneration)** (T7-T8, female, 76 years):
Lateral radiograph *Midsagittal cut*

Grade 3 (severe degeneration) (T9-T10, female, 86 years):
Lateral radiograph *Midsagittal cut*

sclerosis determination. While prioritisation and reduction of measurement parameters generally increase the potential risk of reduced objectiveness, the interobserver reliability showed overall slightly higher agreement values compared to the role models for cervical and lumbar intervertebral disc degeneration grading [10, 11]. This might be explained by both the relatively high agreement values for the variable "Height loss" and the relatively low agreement values for the variable "Diffuse sclerosis" in these studies, which resulted in substantial and almost perfect agreement values, but overall lower agreement values compared to the present study. Moreover, the focus on the evaluation of the two potential anterior osteophytes in the lateral radiograph might explain the overall higher agreement values with regard to the variable "Osteophyte formation", since these osteophytes are considered to be most prominent and easiest to evaluate, especially in the thoracic spine. Another reason for the slightly higher agreement values might be that the interobserver agreement was not tested for one experienced and one unexperienced observer, but for two more experienced

observers in order to increase comparability with previous interobserver reliability studies on thoracic intervertebral disc grading [7, 9]. However, as in the previous studies, one of the two observers was a biomechanical engineer with low medical education, who might be also rated as unexperienced observer. Nevertheless, the effect of unexperienced observers on the objectivity of degeneration grading should be examined in future studies.

Interindividual variations regarding morphology and donor-specific characteristics could have affected the measurements and the reference values on which the new grading system is based. While it is known that intervertebral discs of males exhibit greater degenerative changes compared to females and that intervertebral disc degeneration increases with age [17], the reference values for the height loss measurement were derived from studies which did not specifically include effects of sex and age [12, 13]. Furthermore, physiological diurnal disc height loss could have affected these reference values, while Keller and Nathan predicted lower diurnal height loss per disc in the thoracic compared

to the lumbar spine [18]. Moreover, the sagittal profile might affect the anterior and posterior disc height, especially with regard to radiographs created in upright position. Further limitations of the new grading system might result from the usage of radiographs of dissected specimens for validation and reliability testing, which does not fully reflect the clinical situation where different surrounding tissues superimpose the radiograph, as well as the non-usage of weight-bearing radiographs. However, the new grading system for the thoracic spine was specifically designed to consider potential superimposition and was exclusively based on the evaluation of specific anatomical characteristics and should therefore also be applicable in clinical practice. Nevertheless, the effect of superimposition and weight bearing should be regarded and carefully checked in future examinations. The non-consideration of antero-posterior radiographs, on the other hands, could disregard potential isolated lateral osteophytes, which, however, were not detected in any of the specimens used in this study.

Conclusions

The new radiographic grading system represents a valid method to determine the degree of degeneration of individual thoracic intervertebral discs in a quantitative and therefore more objective manner. This grading scheme can be used either as an additional diagnostic tool for patients showing pain symptoms or for clinical and experimental studies exploring potential relationships between thoracic intervertebral disc degeneration and patient-reported outcome measures or effects on the biomechanics of the thoracic spine. Potential effects of interindividual variations and the superimposition of anatomical structures should be taken into account, especially with regard to specific morphological as well as patient- and donor-specific properties.

References

1. Teraguchi M, Yoshimura N, Hashizume H, Muraki S, Yamada H, Minamide A, Oka H, Ishimoto Y, Nagata K, Kagotani R, Takiguchi N, Akune T, Kawaguchi H, Nakamura K, Yoshida M (2014) Prevalence and distribution of intervertebral disc degeneration over the entire spine in a population-based cohort: the Wakayama Spine study. Osteoarthr Cartil 22(1):104–110

2. 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

3. Pooni JS, Hukins DW, Harris PF, Hilton RC, Davies KE (1986) Comparison of the structure of human intervertebral discs in the cervical, thoracic and lumbar regions of the spine. Surg Radiol Anat 8(3):175–182

4. Healy AT, Mageswaran P, Lubelski D, Rosenbaum BP, Matheus V, Benzel EC, Mroz TE (2015) Thoracic range of motion, stability, and correlation to imaging-determined degeneration. J Neurosurg Spine 23(2):170–177

5. Liebsch C, Jonas R, Wilke HJ (2020) Thoracic spinal kinematics is affected by the grade of intervertebral disc degeneration, but not by the presence of the ribs: an in vitro study. Spine J 20(3):488–498

6. Rutges JP, Duit RA, Kummer JA, Bekkers JE, Oner FC, Castelein RM, Dhert WJ, Creemers LB (2013) A validated new histological classification for intervertebral disc degeneration. Osteoarthr Cartil 21(12):2039–2047

7. Douvier S, Chapurlat R, Estublier C, Szulc P (2020) Reliability of the assessment of disc degeneration on the lateral DXA scans. Jt Bone Spine 88(3):105123

8. Lane NE, Nevitt MC, Genant HK, Hochberg MC (1993) Reliability of new indices of radiographic osteoarthritis of the hand and hip and lumbar disc degeneration. J Rheumatol 20(11):1911–1918

9. Raininko R, Manninen H, Battié MC, Gibbons LE, Gill K, Fisher LD (1995) Observer variability in the assessment of disc degeneration on magnetic resonance images of the lumbar and thoracic spine. Spine 20(9):1029–1035

10. Wilke HJ, Rohlmann F, Neidlinger-Wilke C, Werner K, Claes L, Kettler A (2006) Validity and interobserver agreement of a new radiographic grading system for intervertebral disc degeneration: part I. Lumbar Spine Eur Spine J 15(6):720–730

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

12. Kunkel ME, Herkommer A, Reinehr M, Bockers TM, Wilke HJ (2011) Morphometric analysis of the relationships between intervertebral disc and vertebral body heights: an anatomical and radiographic study of the human thoracic spine. J Anat 219(3):375–387

13. Panjabi MM, Takata K, Goel V, Federico D, Oxland T, Duranceau J, Krag M (1991) Thoracic human vertebrae quantitative three-dimensional anatomy. Spine 16(8):888–901

14. 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

15. Fleiss JL, Cohen J (1973) The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. Educ Psychol Meas 33(3):613–619

16. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33(1):159–174

17. Goh S, Tan C, Price RI, Edmondston SJ, Song S, Davis S, Singer KP (2000) Influence of age and gender on thoracic vertebral body shape and disc degeneration: an MR investigation of 169 cases. J Anat 197(4):647–657

18. Keller TS, Nathan M (1999) Height change caused by creep in intervertebral discs: a sagittal plane model. J Spinal Disord 12(4):313–324