Evaluation of agreement and correlation of results obtained with MRI-based and macroscopic observation–based grading schemes when used to assess intervertebral disk degeneration in cats

Neringa Alisauskaite DVM
Thomas Bitterli DVM
Patrick R. Kircher DVM
Antonio Pozzi DVM
Guy C. M. Grinwis DVM, PhD
Frank Steffen DVM
Lucas A. Smolders DVM, PhD

Received April 9, 2019.
Accepted August 28, 2019.

From the Neurology Service (Alisauskaite, Steffen) and Departments of Small Animal Surgery (Bitterli, Pozzi, Smolders) and Diagnostic Imaging (Kircher), Vetsuisse Faculty, University of Zurich, 8057 Zurich, Switzerland; and Department of Pathobiology, Faculty of Veterinary Medicine, Utrecht University, 3508 TD Utrecht, Netherlands (Grinwis). Dr. Bitterli's present address is the University of Giessen, 35390 Giessen, Germany.

Address correspondence to Dr. Alisauskaite (nalisauskaite@vetclinics.uzh.ch).

OBJECTIVE
To evaluate agreement in results obtained with an MRI-based grading scheme and a macroscopic observation–based grading scheme when used to assess intervertebral disk (IVD) degeneration in cats.

SAMPLE
241 MRI and 143 macroscopic images of singular IVDs in 44 client-owned cats (40 cadaveric and 4 live).

PROCEDURES
Singular images of IVDs were obtained of live cats admitted for treatment of suspected neurologic disease (MRI images of IVDs) and of cadavers of cats euthanized for reasons unrelated to spinal disease (MRI and macroscopic images of IVDs) at the Small Animal Hospital, Vetsuisse Faculty, Zurich, Switzerland, between January 12, 2015, and October 19, 2015. The IVD images were randomized and evaluated twice by 4 observers for each grading scheme. Inter- and intraobserver reliability for the grading schemes was assessed with Cohen weighted κ analysis. Agreement and correlation between results obtained with the 2 grading schemes were determined with Cohen weighted κ and Spearman correlation coefficient (ρ) analyses, respectively.

RESULTS
Inter- and intraobserver agreement between results was substantial to almost perfect (mean weighted κ, 0.66 to 0.83 and 0.71 to 0.86, respectively) for the MRI-based grading scheme and moderate to substantial (mean weighted κ, 0.42 to 0.80 and 0.65 to 0.79, respectively) for the macroscopic observation–based grading scheme. Between the 2 grading schemes, agreement in results was moderate (mean ± SE weighted κ, 0.56 ± 0.05), and the correlation was strong (ρ = 0.73).

CONCLUSIONS AND CLINICAL RELEVANCE
Results indicated that the MRI-based and macroscopic observation–based grading schemes used in the present study could be used reliably for classifying IVD degeneration in cats. (*Am J Vet Res* 2020;81:309–316)

Degenerative spinal disease, a common ailment in people and dogs,[1,2] is characterized by degeneration of different spinal structures, and IVD degeneration has been most investigated.[3–5] To date, MRI is considered the most reliable diagnostic tool to evaluate IVD degeneration status and to grade IVD degeneration in people[6,7] and dogs.[8,9] Although the gold standard for evaluating the degenerative status of IVDs is postmortem macroscopic and histologic evaluation, for antemortem classification purposes, several MRI-based and gross morphologically based grading schemes for lumbar and cervical IVD degeneration have been described in people.[6,7,10–14] Of these, the MRI-based grading scheme reported by Pfirrmann et al[7] and the macroscopic observation–based grading scheme reported by Thompson et al[10] are most widely used and validated for systematic evaluation of IVD degeneration. More recently, these 2 grading schemes have been successfully adapted and validated for evaluating IVD degeneration in dogs.[8,15] Additionally, both of these schemes are known to have substantial agreement with each other,[15] indicating that MRI, to a large extent, reflects gross pathological changes in IVDs.

ABBREVIATIONS

AF	Annulus fibrosus
EP	End plate
IVD	Intervertebral disk
NP	Nucleus pulposus
O1	Observer 1
O2	Observer 2
O3	Observer 3
O4	Observer 4
VB	Vertebral body

Intervertebral disk degeneration in cats has been investigated less extensively than that in people and dogs, and the incidence of IVD disease in cats is lower,[16–19] although the disease is still common.[20–24] Additionally, studies[20,21,24] dating from the 1960s show that cats seem to have IVD degeneration with morphological characteristics similar to those in affected people and dogs. Although the macroscopic and histologic aspects of IVD degeneration in cats have been described,[20–22,24,25] the condition in cats has not been analyzed systematically with standardized grading schemes.

Therefore, the objectives of the present study were to evaluate the agreement and correlation between an MRI-based grading scheme (modified from that reported by Pfirrmann et al[7]) and a macroscopic observation–based grading scheme (modified from that reported by Thompson et al[10]) when used to assess IVD degeneration in cats. We hypothesized that the results for the 2 grading schemes would have substantial to almost perfect inter- and intraobserver agreement and that there would be almost perfect agreement and a strong, positive correlation between results for the 2 grading schemes.

Materials and Methods

Animals

Client-owned cats admitted to the Small Animal Hospital, Vetsuisse Faculty, Zurich, Switzerland, between January 12, 2015, and October 19, 2015, because of suspected neurologic disease and cadavers of cats euthanized at the same facility for reasons unrelated to spinal disease were eligible for inclusion. Owner consent for inclusion of each cat, including cadaveric cats, was obtained. Living cats were only included in evaluation of the MRI-based grading scheme. Cadaveric cats, all collected and processed within 24 hours after death, were included in evaluation of both grading schemes.

Modified Pfirrmann grading scheme

On the basis of results from a previous study[a] and the grading scheme reported by Pfirrmann et al,[7] we developed and used an MRI-based grading scheme that we referred to as the modified Pfirrmann grading scheme. Similar to the original Pfirrmann grading scheme,[7] we used a grading scale of 1 (healthy) to 5 (end-stage degeneration); however, we modified MRI criteria for grades 3 and 4. Specifically for grade 3, the original criterion of clinically normal to slightly decreased IVD height was modified to clinically normal IVD height, and the original criterion for NP structure that appeared nonhomogeneous and gray was modified to NP structure that appeared heterogeneous and gray, with dorsoventrally elongated disk signal. The remaining MRI criteria for grade 3 were unchanged. For grade 4, the original criterion of intermediate to hypointense IVD signal was modified to hypointense IVD signal; all other criteria for grade 4 remained unchanged.

To obtain images of IVDs for evaluation with the modified Pfirrmann grading scheme, MRI was performed on cadaveric and live cats. A 3-T high-field scanner[b] with a head, neck, and spine coil, combined with a posterior coil,[c] was used, and sagittal T2-weighted spin-echo images (repetition time, 3,000 milliseconds; echo time, 100 milliseconds) obtained from the C1-2 IVD to the L7-S1 IVD were obtained at a slice thickness of 2.5 mm. The images were examined on standard computer screens and with use of commercial software.[d] From these images, an independent observer (TB) not involved in assessment with the grading schemes selected individual cervical, thoracic, and lumbar IVDs that had a broad spectrum of different stages of degeneration. For the IVDs selected, singular median IVD images (with the spinal cord located toward the top of each image) were created, labeled to allow blinded and randomized evaluation, and exported as single images for assessment with the modified Pfirrmann grading scheme.

Four independent observers (O1 [NA], a resident in training for certification by the European College of Veterinary Internal Medicine; O2 [LAS], a postdoctoral scientist in degenerative spinal disease; O3 [FS], a diplomate of the European College of Veterinary Internal Medicine; and O4 [PRK], a diplomate of the European College of Veterinary Diagnostic Imaging) then used the modified Pfirrmann grading scheme to grade each of the individual MRI images of IVDs, which were randomized in a spreadsheet program[e] and provided to the observers, who then accessed the images by hyperlinks in the spreadsheet. Each observer graded all IVDs twice, with an interval of 2 months between the first and second rounds of grading. The observers were familiar with the grading scheme reported by Pfirrmann et al,[7] and during the grading procedure, each observer was provided with the modified Pfirrmann grading scheme and corresponding representative images of IVDs from dogs with various grades of degeneration. For IVDs with > 1 grade applicable under the same criterion, the higher grade was selected because the IVD was considered to have had signs of progressive degeneration.

Modified Thompson grading scheme

On the basis of results from a previous study[a] and the grading scheme reported by Thompson et al,[10] we developed and used a macroscopic observation–based grading scheme that we referred to as the modified Thompson grading scheme. We retained the grading scale of 1 (healthy) to 5 (end-stage degeneration) reported by Thompson et al[10] but modified macroscopic criteria for grades 1 and 2. For grade 1, the original criterion of AF with discrete fibrous lamellae was modified to AF with discrete fibrous lamellae and a clear transition zone. For grade 2, the original criterion of AF with mucinous material between lamellae was modified to AF with mucinous material between lamellae and an unclear transition zone.

Guided by findings on MRI of cadaveric cats, the

same independent observer (TB) who selected images from MRIs for use with the modified Pfirrmann grading scheme also collected singular IVDs in different stages of degeneration and from different regions of the vertebral column (cervical, thoracic, or lumbar) from cadaveric cats for use with the modified Thompson grading scheme. Multiple IVDs could be collected from a cadaver, and MRI images of all IVDs collected were available. Each IVD collected was then sawn in the median plane with a water-cooled, diamond-coated band saw.[f] A high-resolution macroscopic image of each sawed IVD was then taken with a digital camera system[g] combined with sample illumination by a 150X photonic cold light source with a photonic standard ring light.[h] Digital photographs were optimized with product-specific imaging management software,[i] and singular high-definition images were created and labeled to allow a blinded, randomized evaluation.

These macroscopic images of IVDs were graded with the modified Thompson grading scheme according to structure and pathological changes in the NPs, AFs, EPs, and VBs. Grading was performed by 3 of the same observers who graded MRI images of IVDs under the modified Pfirrmann grading scheme; however, the diplomate of the European College of Veterinary Diagnostic Imaging (PRK) who participated in that grading was replaced by a veterinary pathologist (GCMG) experienced in grading IVDs in dogs. The macroscopic images of individual IVDs were randomized in the spreadsheet program with a randomization function[e] and provided to the observers, who then accessed the images by hyperlinks in the spreadsheet. These images were then examined on standard computer screens by use of commercial software.[j] Each observer graded images twice, with an interval of at least 2 months between the first and second rounds of grading. The grade assigned to each IVD was the highest grade applicable for any of the structures assessed, and when > 1 grade was applicable for the same criterion, the higher grade was selected.

Statistical analysis

The inter- and intraobserver agreements were determined with the Cohen weighted κ analysis,[26–28] which calculated the percentage of agreement among the grades assigned, corrected by the possibility that the same grades were assigned by chance. Statistical software[k] and an online calculator[l] were used to calculate the SEs and confidence intervals of the weighted κ values obtained. Agreement in results obtained with the 2 different grading schemes was interpreted as slight (weighted κ, 0 to 0.20), fair (weighted κ, 0.21 to 0.40), moderate (weighted κ, 0.41 to 0.60), substantial (weighted κ, 0.61 to 0.80), or almost perfect (weighted κ, 0.81 to 0.99). Agreement and disagreement between results for observers applying the 2 different grading schemes were expressed as numbers and percentages of IVDs assessed under each grading scheme. To investigate whether IVDs from the cervical, thoracic, and lumbar segments of the vertebral column could be graded with the 2 modified schemes, inter- and intraobserver agreement in grades assigned under each of the schemes was calculated separately for IVDs grouped by vertebral column segment (cervical, thoracic, or lumbar).

To calculate the agreement and correlation of results obtained with the 2 grading schemes, and hence whether the grade of an IVD on MRI (assessed with the modified Pfirrmann grading scheme) corresponded to the macroscopic state (assessed with the modified Thompson grading scheme), the mean grade under each scheme was calculated for each IVD by averaging the grades assigned by the 3 observers (O1, O2, and O3) who graded with both schemes. Once the mean grade under each grading scheme was determined for each IVD, Cohen weighted κ analysis and Spearman rank correlation coefficients (ρ) were used to calculate agreement and correlation, respectively. Values of $P < 0.05$ were considered significant.

Results

Animals

Forty-four cats (40 cadaveric and 4 live) were included, and the mean ± SD age was 13 ± 4.5 years (range, 2 to 17 years; **Table 1**; **Supplementary Table S1**, available at: avmajournals.avma.org/doi/suppl/10.2460/ajvr.81.4.309). There were 24 males (21 castrated and 3 sexually intact) and 17 females (13 spayed and 4 sexually intact). All cats had been evaluated at the Small Animal Hospital, Vetsuisse Fac-

Table 1—Descriptive summary of the IVDs in 44 cats with various stages of IVD degeneration imaged (MRI [n = 241], macroscopic [143], or both) for evaluation with the modified Pfirrmann and modified Thompson grading schemes between January 12, 2015, and October 19, 2015, stratified by vertebral column segment and cat age.

		No. of cervical IVDs		No. of thoracic IVDs		No. of lumbar IVDs	
Cat age (y)	No. of cats	MP	MT	MP	MT	MP	MT
≤ 5	5	9	5	7	4	7	4
> 5 and ≤ 10	12	16	11	23	10	16	11
>10	27	43	27	75	44	45	27

MP = Modified Pfirrmann grading scheme (MRI-based grading scheme). MT = Modified Thompson grading scheme (macroscopic observation–based grading scheme).

ulty, Zurich, Switzerland. The 40 cadaveric cats had been euthanized because of reasons unrelated to spinal disease, and the 4 live cats had been evaluated because of IVD disease (n = 3) or myelopathy (1). Further, the 4 live cats were only included in evaluation of the modified Pfirrmann grading scheme.

Table 2—Inter- and intraobserver agreement and disagreement in results obtained by 4 independent observers during 2 rounds of grading the imaged IVDs described in Table 1 under the modified Pfirrmann (n = 241) and modified Thompson (143) grading schemes.

Comparisons	Mean ± SE κ	No (%) IVD grades in agreement	No (%) IVD grades in disagreement by 1 grade	No (%) IVD grades in disagreement by 2 grades	No (%) IVD grades in disagreement by 3 grades
Modified Pfirrmann					
Interobserver round 1					
O1-O2	0.74 ± 0.03	167 (69.3)	73 (30.3)	1 (0.4)	0 (0)
O1-O3	0.73 ± 0.03	160 (66.4)	80 (33.2)	1 (0.4)	0 (0)
O1-O4	0.69 ± 0.03	151 (62.7)	87 (36.1)	3 (1.2)	0 (0)
O2-O3	0.81 ± 0.02	184 (76.3)	57 (23.7)	0 (0)	0 (0)
O2-O4	0.80 ± 0.03	185 (76.8)	54 (22.4)	2 (0.8)	0 (0)
O3-O4	0.83 ± 0.23	191 (79.3)	49 (20.3)	1 (0.4)	0 (0)
Interobserver round 2					
O1-O2	0.69 ± 0.03	159 (66.0)	76 (31.5)	6 (2.5)	0 (0)
O1-O3	0.73 ± 0.03	161 (66.8)	78 (32.4)	1 (0.4)	1 (0.4)
O1-O4	0.71 ± 0.03	164 (68.0)	69 (28.6)	5 (2.1)	3 (1.2)
O2-O3	0.73 ± 0.03	166 (68.9)	73 (30.3)	2 (0.8)	0 (0)
O2-O4	0.66 ± 0.03	157 (65.1)	73 (30.3)	10 (4.1)	1 (0.4)
O3-O4	0.71 ± 0.03	171 (71.0)	59 (24.5)	7 (2.9)	4 (1.7)
Intraobserver					
O1	0.71 ± 0.03	151 (62.7)	86 (35.7)	4 (1.7)	0 (0)
O2	0.86 ± 0.02	202 (83.8)	39 (16.2)	0 (0)	0 (0)
O3	0.83 ± 0.02	190 (78.8)	51 (21.2)	0 (0)	0 (0)
O4	0.81 ± 0.01	189 (78.4)	49 (20.3)	3 (1.2)	0 (0)
Modified Thompson					
Interobserver round 1					
O1-O2	0.74 ± 0.03	90 (62.9)	51 (35.7)	1 (0.7)	1 (0.7)
O1-O3	0.72 ± 0.03	97 (67.8)	43 (30.1)	3 (2.1)	0 (0)
O1-O4	0.69 ± 0.03	87 (60.8)	53 (37.1)	3 (2.1)	0 (0)
O2-O3	0.81 ± 0.02	82 (57.3)	53 (37.1)	6 (4.2)	2 (1.4)
O2-O4	0.80 ± 0.03	65 (45.5)	75 (52.4)	3 (2.1)	0 (0)
O3-O4	0.80 ± 0.23	83 (58.0)	57 (39.9)	2 (1.4)	1 (0.7)
Interobserver round 2					
O1-O2	0.80 ± 0.03	102 (71.3)	40 (28.0)	1 (0.7)	0 (0)
O1-O3	0.65 ± 0.04	78 (54.5)	54 (37.8)	11 (7.7)	0 (0)
O1-O4	0.57 ± 0.04	64 (44.7)	72 (50.3)	6 (4.2)	1 (0.7)
O2-O3	0.73 ± 0.04	90 (62.9)	50 (35.0)	3 (2.1)	0 (0)
O2-O4	0.53 ± 0.04	60 (42.0)	75 (52.4)	8 (5.6)	1 (0.7)
O3-O4	0.42 ± 0.05	39 (27.2)	86 (60.1)	17 (11.9)	1 (0.7)
Intraobserver					
O1	0.77 ± 0.03	95 (66.4)	47 (32.9)	1 (0.7)	0 (0)
O2	0.79 ± 0.03	102 (71.3)	40 (28.0)	0 (0)	1 (0.7)
O3	0.65 ± 0.04	77 (53.8)	58 (40.6)	8 (5.6)	0 (0)
O4	0.74 ± 0.04	102 (71.9)	39 (27.3)	2 (1.4)	0 (0)

Table 3—Interobserver agreement in results depicted in Table 2 stratified by IVD vertebral column segment and grading round.

		Modified Pfirrmann grading scheme						Modified Thompson grading scheme					
		Round 1			Round 2			Round 1			Round 2		
Vertebral column segment	Comparison	Weighted κ	SE	95% CI	Weighted κ	SE	95% CI	Weighted κ	SE	95% CI	Weighted κ	SE	95% CI
Cervical													
	O1-O2	0.73	0.05	0.63–0.83	0.68	0.06	0.57–0.79	0.70	0.06	0.58–0.82	0.73	0.06	0.61–0.85
	O1-O3	0.72	0.05	0.62–0.82	0.65	0.05	0.54–0.75	0.75	0.06	0.62–0.87	0.63	0.08	0.47–0.79
	O1-O4	0.70	0.05	0.61–0.80	0.69	0.05	0.59–0.79	0.66	0.07	0.52–0.80	0.50	0.07	0.36–0.36
	O2-O3	0.88	0.04	0.81–0.96	0.71	0.05	0.61–0.82	0.63	0.07	0.48–0.77	0.67	0.08	0.52–0.82
	O2-O4	0.82	0.05	0.72–0.91	0.66	0.05	0.55–0.77	0.47	0.08	0.32–0.62	0.51	0.08	0.36–0.67
	O3-O4	0.85	0.04	0.76–0.93	0.73	0.05	0.62–0.84	0.65	0.07	0.52–0.79	0.35	0.09	0.18–0.53
Thoracic													
	O1-O2	0.76	0.04	0.68–0.85	0.71	0.04	0.63–0.80	0.76	0.05	0.86–0.86	0.78	0.04	0.69–0.87
	O1-O3	0.68	0.05	0.58–0.77	0.76	0.04	0.68–0.85	0.76	0.05	0.66–0.87	0.49	0.07	0.36–0.63
	O1-O4	0.68	0.05	0.58–0.77	0.65	0.06	0.53–0.77	0.70	0.06	0.58–0.81	0.56	0.07	0.43–0.70
	O2-O3	0.81	0.04	0.74–0.89	0.71	0.05	0.62–0.80	0.70	0.06	0.57–0.82	0.60	0.07	0.47–0.74
	O2-O4	0.76	0.04	0.69–0.85	0.62	0.06	0.50–0.73	0.70	0.05	0.60–0.81	0.50	0.07	0.35–0.64
	O3-O4	0.85	0.03	0.79–0.92	0.66	0.06	0.53–0.78	0.66	0.07	0.53–0.79	0.43	0.07	0.45–0.95
Lumbar													
	O1-O2	0.70	0.05	0.60–0.80	0.64	0.06	0.52–0.76	0.68	0.07	0.54–0.83	0.78	0.08	0.62–0.93
	O1-O3	0.77	0.04	0.69–0.85	0.72	0.05	0.62–0.81	0.70	0.08	0.55–0.84	0.62	0.09	0.43–0.80
	O1-O4	0.67	0.05	0.57–0.77	0.77	0.05	0.68–0.87	0.64	0.08	0.48–0.81	0.58	0.09	0.40–0.76
	O2-O3	0.70	0.05	0.60–0.80	0.66	0.05	0.56–0.76	0.58	0.08	0.43–0.73	0.75	0.07	0.62–0.88
	O2-O4	0.82	0.05	0.73–0.91	0.63	0.06	0.50–0.75	0.62	0.07	0.48–0.77	0.58	0.09	0.41–0.75
	O3-O4	0.76	0.04	0.67–0.85	0.72	0.05	0.61–0.82	0.62	0.07	0.47–0.76	0.46	0.09	0.28–0.64

CI = Confidence interval.

Modified Pfirrmann grading scheme

There were 241 singular, median plane MRI images of separate IVDs (68 cervical, 105 thoracic, and 68 lumbar) evaluated under the modified Pfirrmann grading scheme, and grades ranged from 1 (healthy) to 5 (end-stage degeneration; **Figures 1 and 2**; **Supplementary Table S2**, available at: avmajournals.avma.org/doi/suppl/10.2460/ajvr.81.4.309). Interobserver agreement ranged from substantial (mean ± SE weighted κ, 0.69 ± 0.02) to almost perfect (mean ± SE weighted κ, 0.83 ± 0.23) for the first grading round and was substantial (mean ± SE weighted κ, 0.66 ± 0.03 to 0.73 ± 0.03) for the second grading round **(Table 2)**. The intraobserver agreement under the modified Pfirrmann grading scheme was substantial to almost perfect (mean ± SE weighted κ, 0.71 ± 0.03 [O1]; 0.86 ± 0.02 [O2]; 0.83 ± 0.02 [O3]; and 0.81 ± 0.01 [O4]). When grading results were considered on the basis of vertebral column segment (cervical, thoracic, or lumbar), substantial to almost perfect agreement was detected between observers, with comparable interobserver agreement for each of the 3 segments in both rounds of grading **(Table 3)**.

Modified Thompson grading scheme

There were 143 singular macroscopic images of IVDs (43 cervical, 58 thoracic, and 42 lumbar) from 35 cadaveric cats evaluated with the modified Thompson grading scheme. The remaining 5 cadaveric cats were not used for macroscopic evaluation because there was no possibility to collect fresh macroscopic samples in the appropriate time span after euthanasia. Grades ranged from 1 (healthy) to 5 (end-stage degeneration; Figure 1). Marked degenerative changes (ie, assigned grades of 4 or 5) of the EPs or VBs were seen relatively infrequently (22/143 [15.3%] and 29/143 [20.3%], respectively). New bone formation was observed on the ventral aspect of VBs in only 23 of the 143 (16.1%) IVD images evaluated with the modified Thompson grading scheme, and 4 cats had bony proliferations or bridges in IVDs that otherwise appeared completely healthy (Figure 2).

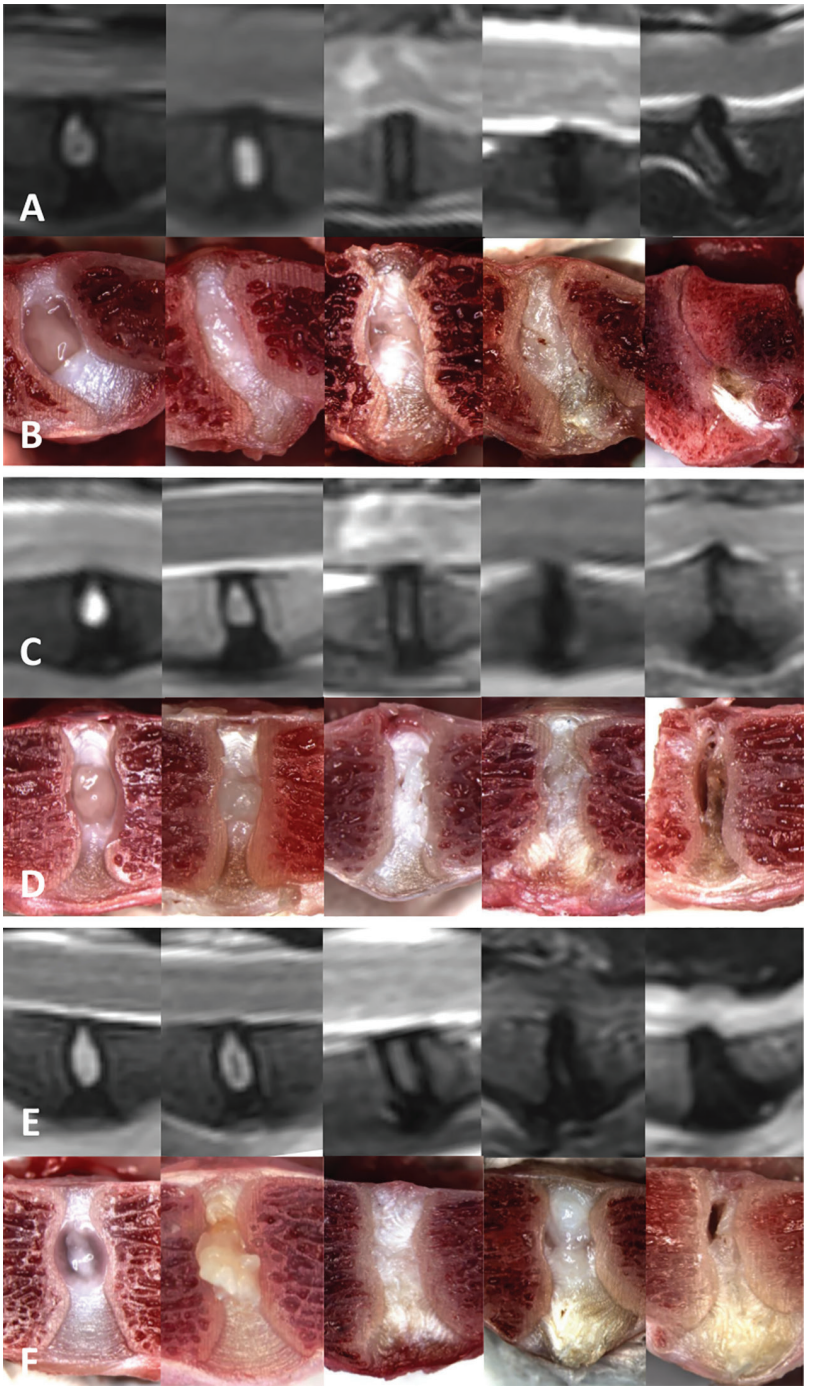

Figure 1—Representative median plane T2-weighted spin-echo MRI (rows A, C, and E) and macroscopic (rows B, D, and F) images of cervical (rows A and B), thoracic (rows C and D), and lumbar (rows E and F) IVDs of 35 cadaveric cats (euthanized for reasons unrelated to spinal disease) showing increasing stages of IVD degeneration as evaluated with the modified Pfirrmann and modified Thompson grading schemes (grades 1 [healthy; far left image of each row] to 5 [end-stage degeneration]; far right image of each row]). In each image, cranial is toward the left, and dorsal is toward the top.

Each observer graded images twice, with an interval of ≥ 2 months between the first and second rounds of grading. One observer (GCMG) performed the second grading round 1 year after the first round and in some instances used magnification to evaluate the EPs

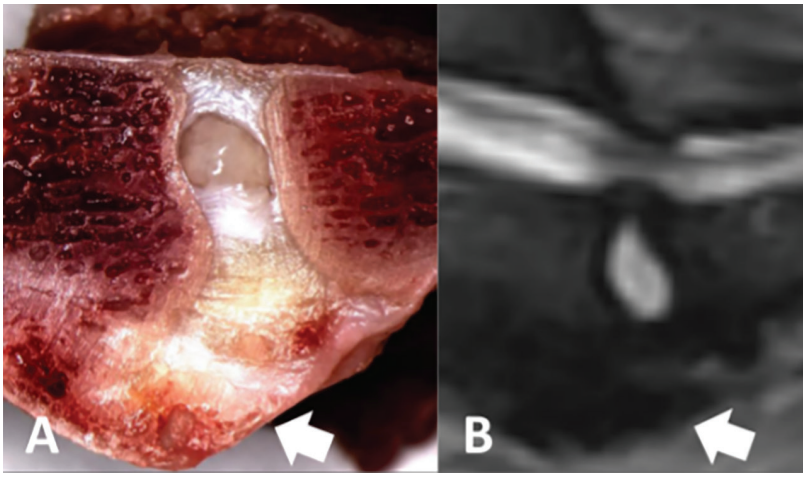

Figure 2—Median plane macroscopic (A) and T2-weighted spin-echo MRI (B) images of the L7-S1 IVD of a 4-kg 6-year-old spayed female European Shorthair cat (euthanized for a reason unrelated spinal disease) showing extensive (> 2 mm) ventral new bone formation (arrows) without any evident signs of degeneration of the NP, AF, or EPs.

during the second round. Interobserver agreement under the modified Thompson grading scheme was substantial (mean ± SE weighted κ, 0.69 ± 0.03 to 0.80 ± 0.23) for the first grading round and ranged from moderate (mean ± SE weighted κ, 0.42 ± 0.04) to substantial (mean ± SE weighted κ, 0.80 ± 0.03) for the second grading round (Table 2). Intraobserver agreement under the modified Thompson grading scheme was substantial (mean ± SE weighted κ, 0.77 ± 0.03 [O1]; 0.79 ± 0.03 [O2]; 0.65 ± 0.04 [O3]; and 0.74 ± 0.04 [O4]). However, O4 consistently assigned higher degeneration grades for the EPs in round 2, compared with round 1, which resulted in higher overall grades assigned to IVDs in round 2, compared with round 1. When grading results were considered on the basis of vertebral column segment (cervical, thoracic, or lumbar), moderate to substantial agreement was detected between observers, with comparable agreement in results for each of the 3 segments in both grading rounds (Table 3).

Agreement and correlation in results from the 2 grading schemes

Agreement between results obtained with the modified Pfirrmann and modified Thompson grading schemes was moderate (mean ± SE weighted κ, 0.56 ± 0.05). In addition, strong positive correlations in results were identified between the modified Pfirrmann and modified Thompson grading schemes overall ($\rho = 0.73$; $P < 0.001$) and for each of the 3 observers ($\rho =$ 0.75 [O1], 0.72 [O2], and 0.77 [O3]) that graded IVDs with both grading schemes.

Discussion

To our knowledge, the present study was the first to report a standardized process for classifying IVD degeneration in cats on the basis of MRI and macroscopic observation. Moderate to almost perfect inter- and intraobserver agreement was detected for both the modified Pfirrmann (mean weighted κ, 0.66 to 0.83 and 0.71 to 0.86, respectively) and the modified Thompson (mean weighted κ, 0.42 to 0.80 and 0.65 to 0.79, respectively) grading schemes. These findings partially supported our hypothesis that the results for the 2 grading schemes would have substantial to almost perfect inter- and intraobserver agreement in that such was identified for the modified Pfirrmann grading scheme; however, our findings for the modified Thompson grading scheme indicated moderate to substantial inter- and intraobserver agreement. Although a strong, positive correlation was identified between results obtained with these 2 grading schemes, the agreement between results was moderate, which was lower than we hypothesized. On the basis of these findings, we rejected our second hypothesis that the results for the 2 grading schemes would have almost perfect agreement and a strong, positive correlation.

The first detailed report[20] describing IVD degeneration in cats divided IVD degeneration into several distinct stages, starting with increased opacity of the NP, followed by immigration of fibrous tissue, separation and distortion of the AF lamellae, and subsequent distortion of the architecture of the disk. Our findings supported that IVD degeneration in cats involves changes of all components of the IVD, including the NP, AF, both EPs, and VBs. Hence, evaluation of these structures with the modified Pfirrmann and modified Thompson grading schemes was applicable to IVDs in cats, which have degenerative changes that are, to a large extent, similar to the pathological changes observed in people and dogs.[10,15] The main alterations we made to the MRI-based grading scheme reported in human medicine by Pfirrmann et al[7] were to provide clarification in descriptions of the structure, shape, and signal intensity of the NP on MRI. Our modifications did not result in inclusion of essentially different characteristics of degeneration. Similarly, we slightly altered the macroscopic observation–based grading scheme reported in human medicine by Thompson et al[10] mostly by further distinguishing grades 1 and 2 by characteristics of the transition zone between the AF and NP, which is clearly defined and described in adult cats.[25] Microscopically, this transition zone has been described in cats as basophilic area that is gradually replaced by dense fibrillar tissue as part of the degeneration process,[25] which is in contrast to humans and dogs in which cell replacement with chondroid tissue in this transition zone is observed at a young age.[29,30] Thus, the loss of a clear transition zone may be used as a microscopic indicator of early IVD degeneration in adult cats.

The agreement in results obtained with the modified Pfirrmann and modified Thompson grading schemes in the present study (mean weighted κ, 0.66 to 0.83 [substantial to almost perfect] and 0.42 to 0.80 [moderate to substantial], respectively) was comparable to findings in people evaluated with the Pfirrmann and Thompson grading schemes (κ, 0.74 to 0.81 and 0.76 to 0.88 [substantial to almost perfect], respectively) but lower than finding in dogs (κ, 0.87 to 0.91 [almost perfect] and 0.69 to 0.94 [substantial to almost perfect], respectively).[7,8,10,15] Several explanations for the discrepancy between these studies can be considered. For instance, in the present study, no training period was included, whereas in the studies[8,10] examining IVDs of dogs, a training period for all observers was included, likely resulting in a higher interobserver agreement. In addition, with respect to the modified Pfirrmann grading scheme, high-field MRI, as used in the present study, may depict more subtle changes in degenerating IVDs that can result in a less clear distinction between various IVD degeneration grades.[31,32] Further, the IVDs of cats are smaller than those of most dogs, rendering images of IVDs in cats more susceptible to volume-averaging artifacts and hampering consistent grading between observers.[33]

Although the IVD morphological changes observed macroscopically during evaluation with the modified Thompson grading scheme in the present study were largely similar to IVD degeneration in people and dogs, several clear distinctions could be observed. In contrast to IVD degeneration in people and dogs, changes of the EPs were seen relatively infrequently in the present study. This finding may have been because EPs in cats are relatively thin, making accurate assessment difficult, whereas EPs in people are thicker and therefore may demonstrate more pronounced distortions with advanced IVD degeneration.[3] To determine whether EPs in cats are subject to substantial degenerative changes, histologic evaluation of EPs of cats is necessary.

Similar to EP changes, degenerative changes of the VBs were relatively infrequent. Ventral new bone formation was observed in only 23 of 143 IVDs. Interestingly, in 4 cats, bony proliferations or bridges were observed in IVDs that otherwise appeared completely healthy. Similar observations have also been reported in dogs.[15,34,35] Although IVD degeneration seems to be the most common cause for bony proliferations of the VBs, there are other causes, such as diffuse idiopathic skeletal hyperostosis, which has been reported in 1 cat.[36]

In the present study, a strong, positive correlation (ρ = 0.73) was detected between results obtained with the modified Pfirrmann grading scheme and those obtained with the modified Thompson grading scheme. These results were comparable with results reported[m] in human medicine in which a coefficient of determination of 0.80 was identified between results obtained with the Pfirrmann grading scheme and a modified Thompson grading scheme when used to evaluate transverse macroscopic IVD images. However, our findings for agreement between results obtained with the 2 modified grading schemes used in the present study (mean weighted κ, 0.56) were lower than the agreement (κ, 0.70) reported for a study[15] in dogs for which the original grading schemes described by Pfirrmann et al[7] and Thompson et al[10] were used. One explanation for this difference could have been that the smaller size of IVDs in cats, compared with those in dogs and people, may have resulted in observers underestimating the macroscopic pathological changes.[33] In addition, MRI slice thickness could have contributed to the lower agreement in the present study because images slightly paramedian could have led to underestimation of IVD degeneration. Furthermore, although MRI of IVDs in cats may reveal marked changes, mainly involving loss of signal intensity and distinction between the NP and AF in the initial degenerative stages, such changes may not have been evident macroscopically and thereby may have contributed to lower agreement in the present study.

A limitation of the present study was that relatively few IVDs were classified as grades 4 or 5, which may have resulted in a relatively low power for assessment of IVDs in higher stages of degeneration. Another limitation was that the interobserver agreement under the modified Thompson grading scheme was substantially lower for O4, compared with that of the other 3 observers. Contributing factors to this finding may have been that the interval between the first and second rounds of grading was notably longer for O4, compared with that for the other 3 observers, and that because EPs in cats are relatively thin and hence difficult to evaluate, O4 used magnification to evaluate the EPs in some instances during the second round of grading. However, as mentioned earlier, the clinical importance of EP changes in cats remains a topic of discussion and requires further evaluation on a histopathologic level.

On the basis of our findings, we believe that the modified Pfirrmann and modified Thompson grading schemes used in the present study can be reliably used to classify IVD degeneration in cats. Results of the present study indicated that use of these 2 modified grading schemes allowed for standardized evaluation of IVD degeneration in cats and provided a strong basis for future studies, including evaluations of clinical cases and in-depth investigations on histologic and biochemical levels, of IVD degeneration in cats.

Acknowledgments

The authors declare that there were no conflicts of interest.

Presented in part as a poster at the 30th Annual Symposium of the European Society of Veterinary Neurology and European College of Veterinary Neurology, Helsinki, Finland, September 2017.

The authors thank Dr. Christian W. A. Pfirrmann for consultation regarding the interpretation of results from the present study.

Footnotes

a. Bitterli T, Ettinger L, Pozzi A, et al. Investigation and grading of intervertebral disc degeneration in the cat by way of magnetic resonance imaging and macroscopic evaluation (abstr). *Vet Comp Orthop Traumatol* 2016;29:A18–A19.
b. Philips Ingenia 3T, Philips Healthcare, Hamburg, Germany.
c. Philips dStream HeadNeckSpine and FlexCoverage posterior coils, Philips Healthcare, Hamburg, Germany.
d. Osirix, Pixmeo SARL, Bern, Switzerland.
e. Microsoft Excel 2016, Microsoft Corp, Redmond, Wash.
f. EXAKT, Exakt Advanced Technologies GmbH, Norderstadt, Germany.
g. Leica DC 480, Leica Microsystems, Wetzlar, Germany.
h. Leica RL 30120201, Leica Microsystems, Wetzlar, Germany.
i. Leica IM1000, Leica Microsystems, Wetzlar, Germany.
j. Preview, MacOS, Apple Inc, Cupertino, Calif.
k. R: a language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. Available at: www.R-project.org. Accessed Dec 14, 2015.
l. Kappa as a measure of concordance in categorical sorting, Lowry R, Avon, Conn. Available at: vassarstats.net/kappa.html. Accessed Jan 27, 2018.
m. Pedowitz DI, Auerbach JD, Gibson BW, et al. Correlation of MRI and gross morphologic grading of lumbar intervertebral discs (abstr). *Trans Orthop Res Soc* 2005:378.

References

1. Resnick D. Degenerative diseases of the vertebral column. *Radiology* 1985;156:3–14.
2. Bergknut N, Smolders LA, Grinwis GCM, et al. Intervertebral disc degeneration in the dog. Part 1: anatomy and physiology of the intervertebral disc and characteristics of intervertebral disc degeneration. *Vet J* 2013;195:282–291.
3. Bergknut N, Rutges JPH, Kranenburg H-J, et al. The dog as an animal model for intervertebral disc degeneration? *Spine (Phila Pa 1976)* 2012;37:351–358.
4. Hansen T, Smolders LA, Tryfonidou MA, et al. The myth of fibroid degeneration in the canine intervertebral disc: a histopathological comparison of intervertebral disc degeneration in chondrodystrophic and nonchondrodystrophic dogs. *Vet Pathol* 2017;54:945–952.
5. Roberts S, Evans H, Trivedi J, et al. Histology and pathology of the human intervertebral disc. *J Bone Joint Surg Am* 2006;88(suppl 2):10–14.
6. Kettler A, Wilke HJ. Review of existing grading systems for cervical or lumbar disc and facet joint degeneration. *Eur Spine J* 2006;15:705–718.
7. Pfirrmann CW, Metzdorf A, Zanetti M, et al. Magnetic resonance classification of lumbar intervertebral disc degeneration. *Spine (Phila Pa 1976)* 26:1873–1878.
8. Bergknut N, Auriemma E, Wijsman S, et al. Evaluation of intervertebral disk degeneration in chondrodystrophic and nonchondrodystrophic dogs by use of Pfirrmann grading of images obtained with low-field magnetic resonance imaging. *Am J Vet Res* 2011;72:893–898.
9. Kärkkäinen M, Punto L, Tulamo R. Magnetic resonance imaging of canine degenerative lumbar spine diseases. *Vet Radiol Ultrasound* 1993;34:399–404.
10. Thompson JP, Pearce RH, Schechter MT, et al. Preliminary evaluation of a scheme for grading the gross morphology of the human intervertebral disc. *Spine (Phila Pa 1976)* 1990;15:411–415.
11. Viikari-Juntura E, Raininko R, Videman T PL. Evaluation of cervical disc degeneration with ultralow field MRI and discography. An experimental study on cadavers. *Spine (Phila Pa 1976)* 1989;14:616–619.
12. Silberstein CE. The evolution of degenerative changes in the cervical spine and an investigation into the "joints of Luschka." *Clin Orthop Relat Res* 1965;40:184–204.
13. Yu S, Haughton VM, Sether LA, et al. Criteria for classifying normal and degenerated lumbar intervertebral disks. *Radiology* 1989;170:523–526.
14. Brant-Zawadzki MN, Jensen MC, Obuchowski N, et al. Interobserver and intraobserver variability in interpretation of lumbar disc abnormalities. A comparison of two nomenclatures. *Spine (Phila Pa 1976)* 1995;20:1257–1263.
15. Bergknut N, Grinwis G, Pickee E, et al. Reliability of macroscopic grading of intervertebral disk degeneration in dogs by use of the Thompson system and comparison with low-field magnetic resonance imaging findings. *Am J Vet Res* 2011;72:899–904.
16. Delisser PJ, Burton NJ. What Is Your Diagnosis? 3-year-old neutered male Burmese cat with a 3-month history of nonspecific hind limb gait abnormalities. *J Am Vet Med Assoc* 2012;240:1289–1290.
17. Danielski A, Bertran J, Fitzpatrick N. Management of degenerative lumbosacral disease in cats by dorsal laminectomy and lumbosacral stabilization. *Vet Comp Orthop Traumatol* 2013;26:69–75.
18. De Decker S, Warner A-S, Volk HA. Prevalence and breed predisposition for thoracolumbar intervertebral disc disease in cats. *J Feline Med Surg* 2017;19:419–423.
19. Harris JE, Dhupa S. Lumbosacral intervertebral disk disease in six cats. *J Am Anim Hosp Assoc* 2008;44:109–115.
20. King AS, Smith RN. Degeneration of the intervertebral disc in the cat. *Acta Orthop Scand* 1964;34:139–158.
21. Butler WF. Histological age changes in the ruptured intervertebral disc of the cat. *Res Vet Sci* 1968;9:130–135.
22. King A, Smith R. Disc protrusions in the cat: distribution of dorsal protrusions along the vertebral column. *Vet Rec* 1960;72:335–337.
23. King A, Smith R. Disc protrusion in the cat: age incidence of dorsal protrusions. *Vet Rec* 1960;72:381–382.
24. King A, Smith R. Protrusion of the intervertebral disc in the cat. *Vet Rec* 1958;70:509–515.
25. Butler WF, Smith RN. Age changes in the nucleus pulposus of the non-ruptured intervertebral disc of the cat. *Res Vet Sci* 1967;8:151–156.
26. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–174.
27. Landis JR, Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics* 1977;33:363–374.
28. Koch GG, Landis JR, Freeman JL, et al. A general methodology for the analysis of experiments with repeated measurement of categorical data. *Biometrics* 1977;33:133–158.
29. Bergknut N, Meij BP, Hagman R, et al. Intervertebral disc disease in dog—part 1: a new histological grading scheme for classification of intervertebral disc degeneration in dogs. *Vet J* 2013;195:156–163.
30. Hansen HJ. A pathologic-anatomic study on disc degeneration in dog, with special reference to the so-called enchondrosis intervertebralis. *Acta Orthop Scand* 1952;23(suppl 11):1–117.
31. Wijayathunga VN, Ridgway JP, Ingham E, et al. A nondestructive method to distinguish the internal constituent architecture of the intervertebral discs using 9.4 Tesla magnetic resonance imaging. *Spine (Phila Pa 1976)* 2015;40:E1315–E1322.
32. Welsch GH, Trattnig S, Paternostro-Sluga T, et al. Parametric T2 and T2* mapping techniques to visualize intervertebral disc degeneration in patients with low back pain: initial results on the clinical use of 3.0 Tesla MRI. *Skeletal Radiol* 2011;40:543–551.
33. Gavin PR. Artifacts. In: Gavin PR, Bagley RS, eds. *Practical small animal MRI.* Ames, Iowa: Wiley-Blackwell, 2009;10–20.
34. Seiler G, Häni H, Scheidegger J, et al. Staging of lumbar intervertebral disc degeneration in nonchondrodystrophic dogs using low-field magnetic resonance imaging. *Vet Radiol Ultrasound* 2003;44:179–184.
35. Wright JA. Spondylosis deformans of the lumbo-sacral joint in dogs. *J Small Anim Pract* 1980;21:45–58.
36. Bossens K, Bhatti S, Van Soens I, et al. Diffuse idiopathic skeletal hyperostosis of the spine in a nine-year-old cat. *J Small Anim Pract* 2016;57:33–35.