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

Niklas Bergknut, DVM, PhD; Guy Grinwis, DVM, PhD; Emile Pickee, BSc; Edoardo Auriemma, DVM; Anne-Sofie Lagerstedt, DVM, PhD; Ragnvi Hagman, DVM, PhD; Herman A. W. Hazewinkel, DVM, PhD; Björn P. Meij, DVM, PhD

Objective—To evaluate the reliability of the Thompson system for use in grading the gross pathological changes of intervertebral disk (IVD) degeneration in dogs and to investigate the agreement between gross pathological findings and low-field (0.2-T) magnetic resonance imaging (MRI) findings.

Sample—Vertebral columns from cadavers of 19 dogs of various ages, breeds, and origins.

Procedures—182 intervertebral segments were collected from 19 canine cadavers. Sagittal T2-weighted MRI of the T11 through S1 portion of the vertebral column was performed within 24 hours after the dogs were euthanized. The vertebral columns were subsequently divided in the midsagittal plane, and high-resolution photographs were obtained of each intervertebral segment (end plate–disk–end plate). The MRI images and photographs were graded separately in a blinded manner by 4 observers who used both Pfirrmann and Thompson grading criteria.

Results—The interobserver agreement for Thompson scores ranged from 0.76 to 0.88, and the intraobserver agreement ranged from 0.88 to 0.94 (Cohen weighted κ analysis). Agreement between scores for the Pfirrmann and Thompson grading criteria was κ = 0.70.

Conclusions and Clinical Relevance—Grading of IVD degeneration in dogs by use of the Thompson system resulted in high interobserver and intraobserver agreement, and scores for the Thompson system had substantial agreement with low-field MRI findings graded by use of the Pfirrmann system. This suggested that low-field MRI can be used to diagnose IVD degeneration in dogs. (Am J Vet Res 2011;72:899–904)
sidered to include degeneration attributable to genetic or pathological reasons as well as to typical physiologic aging. It is difficult to distinguish between pathological and physiologic age-related changes in IVDs. Moreover, physiologically normal age-related changes (so-called senile remodeling) can ultimately lead to pathological conditions in IVDs.

Many similarities exist in IVD degeneration between humans and dogs. For example, the clinical signs, diagnostic methods and treatments used, loading patterns, and macroscopic and microscopic appearance are similar. However, although IVD degeneration is generally of the chondroid type in chondrodystrophic humans and dogs as well as achondroplastic humans, IVD degeneration is generally of the fibrous type in nonchondrodystrophic humans and dogs.

In human medicine, there are established systems for scoring the severity of disk degeneration on the basis of histologic examination MRI, and gross pathological changes. The most commonly used criterion-referenced standard for IVD degeneration in humans is the 5-category grading scheme for gross pathological changes developed by Thompson et al., although grading of disk degeneration on MRI in accordance with the system developed by Pfirrmann et al. for evaluation of lumbar IVDs is also used extensively. Such scoring systems for IVD degeneration are not typically used in veterinary medicine. A grading system has been proposed for MRI images of lumbar IVDs by use of histologic examination as the criterion-referenced standard. In a study, investigators described MRI findings and the gross morphological appearance of degenerated canine IVDs, and investigators in another study reported the correlation between disk protrusion detected by use of MRI and computed tomography and surgical findings.

To the authors' knowledge, the Thompson system has not been used to grade IVD degeneration in dogs, although there is no other scoring system available for evaluating the gross pathological changes of IVD degeneration in dogs. The purpose of the study reported here was to evaluate the reliability of the Thompson system for use in grading gross pathological changes of IVD degeneration in dogs and to investigate the agreement between gross pathological findings and low-field MRI findings.

**Materials and Methods**

**Sample**—The vertebral columns from 19 cadavers of dogs of various breeds and ages were included in the study. Dogs < 1 year old were not included because they rarely have disk degeneration. There were 5 Beagles, 2 Bouvier des Flandres, 3 Foxhounds, 3 Kerry Beagles, 1 Flat-Coated Retriever, 1 Welsh Terrier, and 4 mixed-breed dogs. The dogs (12 females and 7 males) ranged from 1 to 16 years of age and weighed 9 to 44 kg. The dogs were research dogs or client-owned dogs (with permission from the owners) that had been examined at the Utrecht University Clinic for Companion Animals. All dogs had been euthanized for reasons unrelated to the study reported here.

**Collection and processing of the vertebral columns**—The T11 through S1 vertebral segments were dissected from each of the 19 canine cadavers at a pathobiology laboratory within 24 hours after the dogs were euthanized. Five of the vertebral segments were cut at T12 through L1 instead of at T11, which resulted in 182 intervertebral segments that were used in the study. The vertebral segments were wrapped in moist towels, and sagittal T2-weighted MRI images were obtained within 1 hour after segment dissection by use of a low-field (0.2-T) open magnet with a 16-cm-diameter multipurpose flexible coil. It was necessary to dissect the vertebral segments before MRI because it was not possible to transport entire cadavers to the MRI facility. This might have led to a slight decrease in MRI resolution because the coil diameter was slightly larger than the diameter of the vertebral bodies, which can lead to a decrease in the signal-to-noise ratio. The MRI process included obtaining sagittal T2-weighted spin echo images (repetition time, 3,835 to 4,450 milliseconds; echo time, 117 milliseconds). The obtained MRI images had a slice thickness of 3 mm, and the slice that best depicted the IVD was selected for grading of IVD degeneration by use of the method of Pfirrmann et al.

After MRI, the vertebrae were sectioned through the midline (sagittal plane) with a water-cooled belt saw, and the midsagittal cut surfaces of the intervertebral segments were cleaned to remove debris. Vertebrae were placed on a table. Multiple lights were used to illuminate the vertebrae from different directions, and high-resolution photographs were obtained by use of a digital single-lens reflex camera equipped with a macro flash. A transparent ruler was placed on the cut surface as a scale marker. The digital photographs were cropped by use of photographic management software so that a single intervertebral segment was visible in each photograph. This was performed to prevent the observers who graded IVD degeneration from having knowledge about the source of the intervertebral segment (ie, location in the vertebral column) and also to prevent comparison with adjacent intervertebral segments. The photographs were used for the gross morphology grading of IVD degeneration in accordance with the method of Thompson et al.

**Pfirrmann grading of MRI images**—The Pfirrmann system for grading IVD degeneration in human lumbar disks on midsagittal MRI images is a reliable method for use in grading IVDs of dogs. The grading system was based on 4 criteria (signal intensity, disk structure, disk height, and distinction between the nucleus and annulus; Appendix 1). The MRI images were converted to a digital format (ie, JPEG) and placed in a spreadsheet program to prevent identification of the images by the observers; observers accessed the images via hyperlinks to the spreadsheet program. Grading was performed on T2-weighted MRI images; Pfirrmann scores ranged from grade 1 (healthy IVD) to 5 (end-stage IVD degeneration). Four observers (a veterinary medical student, a student in a PhD program, a board-certified veterinary surgeon, and a board-certified veterinary radiologist)
separately graded the MRI images: none of the observers were aware of the source of the IVDs.

Thompson grading of photographs—The grading system described by Thompson et al.21 is a 5-category method for assessing the gross morphology of mid-sagittal sections of human lumbar IVDs. Pathological changes of the nucleus pulposus, annulus fibrosus, end plates, and periphery of the vertebral body (Appendix 2) were graded by the same 4 observers who performed the Pfirrmann grading of MRI images, except that the board-certified radiologist was replaced by a board-certified veterinary pathologist. The photographs were provided to the observers in random order and in duplicate in a spreadsheet program; they were accessed via hyperlinks to the spreadsheet program.

Statistical analysis—Interobserver and intraobserver reliability of the Thompson grading system were analyzed by use of Cohen weighted κ analysis, which calculated the percentage of agreement among the grades assigned to an image corrected by the chance that the same grade was assigned by chance. The weighted κ analysis provides more weight to grades that differ from each other by only 1 unit and less weight to grades that differ by >1 unit. Agreement was interpreted as follows: slight (κ, 0 to 0.20), fair (κ, 0.21 to 0.40), moderate (κ, 0.41 to 0.60), substantial (κ, 0.61 to 0.80), and almost perfect (κ, 0.81 to 1.00).30–32 To calculate interobserver reliability (the frequency of agreement between the 4 observers), grades assigned by the observer with the highest intraobserver agreement (the veterinary medical student, κ = 0.94) were selected for comparison with the grades assigned by the 3 other observers.

After evaluation of interobserver and intraobserver reliability, the mean of the values assigned by the 4 observers for all IVDs investigated (Pfirrmann grades for the MRI images and Thompson grades for the photographs) was calculated to enable analysis of the agreement between the Thompson grades and the degree of degeneration on MRI images. In cases in which there was a deviation of >1 grade between the 4 observers, the image was reviewed again simultaneously by all 4 observers and a consensus decision was made. Agreement between the mean Pfirrmann and Thompson grades for each intervertebral segment was then analyzed by use of Cohen weighted κ analysis.

Results

Descriptive results—The frequency of IVD degeneration on the MRI images (Figure 1) and photographs (Figure 2) decreased in a similar pattern from grade 1 to 5. By use of the Pfirrmann scale, IVD degeneration in MRI images was assigned a grade of 1 for 71 of 182 (39.0%) intervertebral segments, a grade of 2 for 61 of 182 (33.5%) intervertebral segments, a grade of 3 for 41 of 182 (22.5%) intervertebral segments, a grade of 4 for 6 of 182 (3.3%) intervertebral segments, and a grade of 5 for 3 of 182 (1.6%) intervertebral segments. By use of the Thompson scale, IVD degeneration in photographs was assigned a grade of 1 for 61 of 182 (33.5%) intervertebral segments, a grade of 2 for 74 of 182 (40.7%) intervertebral segments, a grade of 3 for 27 of 182 (14.8%) intervertebral segments, a grade of 4 for 16 of 182 (8.8%) intervertebral segments, and a grade of 5 for 4 of 182 (2.2%) intervertebral segments.

Interobserver reliability of the Thompson system—Interobserver reliability was almost perfect for all observers, with mean ± SE κ scores ranging from 0.94 ± 0.07 to 0.88 ± 0.07. In most cases (138/182 [75.8%] to 162/182 [89.0%]), the 2 photographs of the same intervertebral segment received the same grade. Differences between the grades for duplicate photographs were generally only 1 grade, although there was a difference of 2 grades for 2 intervertebral segments and a difference of 3 grades for 1 intervertebral segment.

Interobserver reliability of the Thompson system—Interobserver reliability was almost perfect (mean ± SE κ, 0.88 ± 0.07 to 0.83 ± 0.07) for 5 of 6 comparisons between observers and was substantial.
Agreement between grades for the Pfirrmann system and Thompson system—Agreement was substantial (κ = 0.70) between mean grades for the Thompson system and mean grades for the Pfirrmann system. Agreement between grades for the Thompson and Pfirrmann systems was lower for the segments with IVD degeneration grades 4 and 5 than for the segments with IVD degeneration grades 1 to 3; however, there were fewer segments with severe (grades 4 and 5) IVD degeneration than with less severe (grades 1 to 3) IVD degeneration.

**Discussion**

Most diagnostic methods and treatments used in veterinary medicine have been adapted from methods and treatments used in human medicine. However, this is not always possible because of anatomic or physiologic differences between humans and other animal species. For this reason, new diagnostic methods, treatments, or surgical techniques need to be validated in the species in which they are intended for use before being implemented in veterinary practice. In the study reported here, we investigated whether the Thompson system used to grade IVD degeneration in humans could be applied to dogs. We found that the intraobserver and interobserver agreement for grading IVD degeneration by use of the Thompson system was almost perfect, which indicated that the Thompson system can be used to reliably grade IVD degeneration in cadavers of dogs.

To develop treatments for dogs with early IVD degeneration, diagnostic imaging methods with a high sensitivity and specificity for detecting early signs of disease are needed. This requires a reliable pathologic-anatomic criterion-referenced standard that is reproducible and can be used to distinguish among stages of disease. The Thompson system fulfills these criteria and is the current criterion-referenced standard for research in IVD degeneration in humans. Thus, we chose the Thompson system for use in grading IVD degeneration in dogs. Other methods for determining the degree of disk degeneration include histologic examination and measurement of the glycosaminoglycan content of the nucleus. Although these methods have advantages, they lack the simplicity and overview of the Thompson system.

Although the Thompson grading system worked well when applied to IVDs of dogs, we detected interspecies differences in the development of vertebral lesions, with ventral and cranial spondylosis being common in dogs. In humans, spondylosis is generally associated with advanced stages of IVD degeneration, whereas spondylosis was encountered in only a few affected (and even some healthy) IVDs in the present study. This has been reported in the veterinary literature. The presence of spondylosis in vertebral segments with healthy or mildly degenerated IVDs might have biased the observers in assigning grades that were too high. According to the Thompson system, osteophytes should only be present in severe degeneration (grades 3 to 5), and osteophytes > 2 mm should only be present in the severest degeneration (grade 5). Thus, detection of spondylosis would be considered clinically relevant to IVD degeneration only when spondylosis is present in severely degenerated IVDs. Spondylosis in dogs is associated with type II disk herniation rather than with type I herniation, which supports the hypothesis that fibrous metaplasia of the IVD in dogs resembles IVD degeneration in humans.

By use of the Thompson system, there was an intraobserver difference of 2 grades for 2 intervertebral segments and of 3 grades in another intervertebral segment. We cannot explain these findings because these segments did not share features that distinguished them from the other intervertebral segments. Therefore, it is not clear whether there were clerical errors or whether the observer indeed perceived the degeneration as being markedly different between the 2 grading sessions.

Several factors could have influenced the Pfirrmann grading of IVD degeneration on MRI images of dogs. First, the dogs were of various ages, shapes, and sizes. Second, resolution of the disks is markedly lower in small-breed dogs than in large-breed dogs when low-field MRI is used because of the smaller size of the IVDs (image resolution is the same for small and large dogs). This could have led to inaccurate grading of the IVDs of small dogs. Third, T2-weighted MRI images were used because they best depict the glycosaminoglycan and water content of the disks, which in turn is correlated with the degree of disk degeneration. This means that the brighter the appearance of the nucleus on a T2-weighted MRI image, the higher the glycosaminoglycan content and the healthier the disk was. The so-called coil effect of the MRI results in a brighter signal in the focus area of the magnetic field than in the adjacent areas of the vertebral column, which could lead to falsely low signals for the nucleus pulposus in IVDs located peripherally in the images and thus falsely high Pfirrmann scores. Moreover, the vertebral columns had been removed before MRI, which would have resulted in slightly decreased MRI resolution because the coil diameter was slightly larger than the diameter of the vertebral bodies, which would lead to a decreased signal-to-noise ratio.

The overall agreement between the mean grades for the Thompson system and the mean grades for the Pfirrmann system was substantial (κ, 0.70). Most IVDs with a Thompson grade of 1 or 2 were assigned a Pfirrmann grade of 1 or 2, and the deviation was ≤ 1 grade. However, for IVDs with a Thompson grade of 3, 4, or 5, the Pfirrmann grade deviated by up to 2 grades, and for IVDs with a Thompson grade of 4 or 5, less than half received the same Pfirrmann grades. Few disks graded by use of the Thompson and Pfirrmann systems had a grade of 4 (16 [8.8%] and 6 [3.3%], respectively) or 5 (4 [2.2%] and 3 [1.6%], respectively). These low frequencies may have contributed to the poor agreement between the systems for grades 4 and 5. Another factor that may have contributed to the low agreement is the intrinsic difference between the 2 grading systems. The Thompson system is based on changes in the nucleus, annulus, end plate, and vertebral body, whereas the Pfirrmann system is based on changes in the

902

AJVR, Vol 72, No. 7, July 2011
nucleus and annulus. Future studies should focus on the agreement between the MRI findings and the macroscopic appearance of high-grade IVD degeneration.

The high amount of agreement between macroscopic changes (graded by use of the Thompson system) and low-field MRI findings (graded by use of the Pfirrmann system) suggests that on the basis of the increasing availability of MRI in veterinary practices, MRI will become an increasingly useful tool for the diagnosis of IVD degeneration in dogs in a clinical setting. However, as previously stated, dogs often have degenerated IVDs without any clinical signs, so a finding of degenerated IVDs on MRI images will have to be substantiated by clinical and neurologic findings to reach an accurate diagnosis.

In the study reported here, we determined that the Thompson system was a reliable method for grading IVD degeneration in dogs, with a high interobserver and intraobserver agreement. Furthermore, there was substantial agreement between macroscopic grades of intervertebral segments assigned in accordance with the Thompson system and grades of low-field MRI images assigned in accordance with the Pfirrmann system, which suggests that evaluation of low-field MRI images can be used to diagnose IVD disease in dogs. The agreement was stronger in the earlier stages of degeneration (grades 1 to 3), which is advantageous because it enables early clinical intervention. However, because dogs may have subclinical IVD degeneration, it is important that clinical and neurologic findings be used to substantiate MRI findings of IVD degeneration to reach an accurate diagnosis.

References

### Appendix 1

Description of the MRI-based grading system used to classify IVDs in dogs.

<table>
<thead>
<tr>
<th>Pfirrmann grade</th>
<th>Structure</th>
<th>Distinction between NP and AF</th>
<th>Signal intensity</th>
<th>Height of IVD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Homogeneous and bright white</td>
<td>Clear</td>
<td>Hyperintense and isointense to CSF</td>
<td>Normal</td>
</tr>
<tr>
<td>2</td>
<td>Nonhomogeneous with or without horizontal bands</td>
<td>Clear</td>
<td>Hyperintense and isointense to CSF</td>
<td>Normal</td>
</tr>
<tr>
<td>3</td>
<td>Nonhomogeneous and gray</td>
<td>Unclear</td>
<td>Intermediate</td>
<td>Normal to slightly decreased</td>
</tr>
<tr>
<td>4</td>
<td>Nonhomogeneous and gray to black</td>
<td>Lost</td>
<td>Intermediate to hypointense</td>
<td>Normal to moderately decreased</td>
</tr>
<tr>
<td>5</td>
<td>Nonhomogeneous and black</td>
<td>Lost</td>
<td>Hypointense</td>
<td>Collapsed disk space</td>
</tr>
</tbody>
</table>

AF = Annulus fibrosus, NP = Nucleus pulposus.

### Appendix 2

Description of the macroscopic grading scheme used to classify gross pathological changes of IVDs in dogs.

<table>
<thead>
<tr>
<th>Thompson grade</th>
<th>Nucleus pulposus</th>
<th>Annulus fibrosus</th>
<th>End plates</th>
<th>Vertebral bodies</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bulging gel</td>
<td>Discrete fibrous lamellae</td>
<td>Hyaline; uniform thickness</td>
<td>Rounded margins</td>
</tr>
<tr>
<td>2</td>
<td>White fibrous tissue peripherally</td>
<td>Mucinous material between lamellae</td>
<td>Irregular thickness</td>
<td>Pointed margins</td>
</tr>
<tr>
<td>3</td>
<td>Consolidated fibrous tissue</td>
<td>Extensive mucinous infiltration; loss of annular-nuclear demarcation</td>
<td>Focal defects in cartilage</td>
<td>Early chondrocytes or osteophytes at margins</td>
</tr>
<tr>
<td>4</td>
<td>Horizontal (vertical) clefts parallel to end plate</td>
<td>Focal disruptions</td>
<td>Fibrocartilage extending from subchondral bone; irregularity and focal sclerosis in subchondral bone</td>
<td>Osteophytes &lt; 2 mm</td>
</tr>
<tr>
<td>5</td>
<td>Clefts extend through nucleus and annulus</td>
<td>Diffuse sclerosis</td>
<td>Diffuse sclerosis</td>
<td>Osteophytes &gt; 2 mm</td>
</tr>
</tbody>
</table>