Interobserver agreement and diagnostic accuracy of brain magnetic resonance imaging in dogs

Mylène-Kim Leclerc, DMV, DACVIM; Marc-André d'Anjou, DMV, DACVR; Laurent Blond, DMV, MS, DACVR; Éric Norman Carmel, DMV; Ruth Dennis, DMV, VetMB, MA; Susan L. Kraft, DVM, PhD, DACVR, DACVR-RO; Andrea R. Matthews, DMV, DACVR; Joane M. Parent, DMV, MVetSc, DACVIM

Objective—To evaluate interobserver agreement and diagnostic accuracy of brain MRI in dogs.

Design—Evaluation study.

Animals—44 dogs.

Procedures—5 board-certified veterinary radiologists with variable MRI experience interpreted transverse T2-weighted (T2w), T2w fluid-attenuated inversion recovery (FLAIR), and T1-weighted–FLAIR; transverse, sagittal, and dorsal T2w; and T1-weighted–FLAIR postcontrast brain sequences (1.5 T). Several imaging parameters were scored, including the following: lesion (present or absent), lesion characteristics (axial localization, mass effect, edema, hemorrhage, and cavitation), contrast enhancement characteristics, and most likely diagnosis (normal, neoplastic, inflammatory, vascular, metabolic or toxic, or other). Magnetic resonance imaging diagnoses were determined initially without patient information and then repeated, providing history and signalment. For all cases and readers, MRI diagnoses were compared with final diagnoses established with results from histologic examination (when available) or with other pertinent clinical data (CSF analysis, clinical response to treatment, or MRI follow-up). Magnetic resonance scores were compared between examiners with κ statistics.

Results—Reading agreement was substantial to almost perfect ($0.64 < \kappa < 0.86$) when identifying a brain lesion on MRI; fair to moderate ($0.14 < \kappa < 0.60$) when interpreting hemorrhage, edema, and pattern of contrast enhancement; fair to substantial ($0.22 < \kappa < 0.74$) for dural tail sign and categorization of margins of enhancement; and moderate to substantial ($0.40 < \kappa < 0.78$) for axial localization, presence of mass effect, cavitation, intensity, and distribution of enhancement. Interobserver agreement was moderate to substantial for categories of diagnosis ($0.56 < \kappa < 0.69$), and agreement with the final diagnosis was substantial regardless of whether patient information was ($0.65 < \kappa < 0.76$) or was not ($0.65 < \kappa < 0.68$) provided.

Conclusions and Clinical Relevance—The present study found that whereas some MRI features such as edema and hemorrhage were interpreted less consistently, radiologists were reasonably constant and accurate when providing diagnoses. (*J Am Vet Med Assoc* 2013;242:1688–1695)

Magnetic resonance imaging is now routinely used to identify brain lesions noninvasively, establish prognoses, and develop treatment protocols; however, little is known about the diagnostic accuracy of this modality and the potential disagreement that may exist between readers. Numerous studies have described and established the prevalence of variable MRI characteristics in various types of brain disease in dogs, such as noninfectious meningoencephalitis,[1–7] brain tumors,[2,8–19] vascular infarcts,[20–26] and infectious meningoencephalitis.[2,27,28] Nonetheless, these characteristics were most often described by a single observer or by consensus, which may not appropriately reflect clinical practice. Indeed, a radiologist providing a report remotely without close interaction with the referring veterinarian may not interpret

Abbreviations

CHUV-UM	Centre hospitalier universitaire vétérinaire of the Université de Montréal
FLAIR	Fluid-attenuated inversion recovery
FSE	Fast spin echo
GRE	Gradient-recalled echo
T1w	T1-weighted
T2w	T2-weighted

From the Département de sciences cliniques, Faculté de médecine vétérinaire, Université de Montréal, Saint-Hyacinthe, QC J2S 7C6, Canada (Leclerc, d'Anjou, Blond, Carmel, Parent); The Animal Health Trust, Landwades Park, Kentford, Newmarket, Suffolk, CB8 7UU, England (Dennis); the Department of Environmental and Radiological Health Sciences, Veterinary Teaching Hospital, Colorado State University, Fort Collins, CO 80523 (Kraft); and the Department of Small Animal Clinical Sciences, College of Veterinary Medicine, University of Tennessee, Knoxville, TN 37996 (Matthews). Dr. Leclerc's present address is Hôpital Vétérinaire Rive Sud, 7415, Taschereau Ouest, Brossard, QC J4Y 1A2, Canada.

This study was performed at the Centre hospitalier universitaire vétérinaire, Saint-Hyacinthe, QC, Canada.

Presented in part as an abstract at the American College of Veterinary Internal Medicine Forum, Anaheim, Calif, June 2010.

The authors thank Guy Beauchamp for statistical analysis.

Address correspondence to Dr. Leclerc (mylene-kim.leclerc@hvrs.com).

images the same way as would a radiologist involved in a research project on a specific disease or a group of radiologists interpreting images in consensus.

Several MRI features, such as axial lesion localization, cerebral edema, mass effect, contrast enhancement, and presence of dural tail sign, among others, are considered when describing brain disease. As these criteria can vary among disease types, their recognition and grading can influence the imaging diagnosis. For instance, a strongly contrast-enhancing, extra-axial lesion associated with a dural tail sign is indicative of meningioma in dogs,[29] whereas contrast enhancement secondary to reperfusion should be faint or only peripheral in chronic ischemic infarction.[20,23,25] Most of these features have been extensively described, but their gradation schemes are not standardized. Hence, the variability among radiologists in identifying and grading these imaging characteristics is potentially high.

Besides imaging characteristics, knowledge of a patient's signalment has also been reported to be an important factor when interpreting brain MRI images.[30] Indeed, certain brain anomalies, such as hydrocephalus and arachnoid cysts of the quadrigeminal cistern, can be observed incidentally on MRI images in some breeds without associated clinical signs.[30–34] More importantly, distinct disorders may present similar imaging characteristics, therefore requiring a strong understanding of the patient's clinical signs and signalment for better disease discrimination.[9,35] To our knowledge, the impact of a patient's signalment and clinical history in establishing a diagnosis on the basis of brain MRI images has not been determined in veterinary medicine.

The purpose of the study reported here was to evaluate interobserver agreement and diagnostic accuracy in the interpretation of canine brain MRI images. Our hypotheses were that the interpretation of several criteria such as brain edema and hemorrhage would vary among radiologists, the overall interobserver agreement for final diagnosis would be substantial to excellent ($\kappa \geq 0.61$), and providing patient history and signalment to the radiologist would influence the diagnosis and would increase the level of diagnostic confidence.

Materials and Methods

Criteria for patient selection—Medical records of dogs that had brain MRI performed at the CHUV-UM between January 2007 and May 2009 were reviewed. Only patients evaluated at least once by a neurologist prior to MRI and in which intracranial disease was neurologically localized were considered. Images were reviewed by 2 of the investigators (MKL and ENC) who were not involved in the clinical reading to confirm that image sequences were consistently acquired and were of optimal quality. From these patients, only dogs that fulfilled at least one of the following criteria were included in the study: had histologic analysis confirming intracranial pathological changes or normality, had available additional MRI sequences (eg, diffusion) and follow-up MRI validating initial diagnoses and at least 2 follow-up examinations by a neurologist (eg, infarction), or had CSF analysis and at least 2 follow-up examinations by a neurologist confirming response to treatment for the presumed diagnosis (eg, inflammatory disease).

MRI—All images were acquired with a 1.5-T system[a] and a dedicated knee[b] or spine[c] coil depending on the size of the patient. Several additional sequences (eg, T2*-GRE and diffusion-weighted imaging) and imaging planes were obtained in most dogs, but only the following sequences were provided to the readers: transverse, dorsal, and sagittal T2w-FSE; transverse T2w-FLAIR; transverse T1w-FLAIR pre- and postcontrast (5 minutes following IV injection of 0.2 mL of gadobenate dimeglumine[d]/kg [0.09 mL/lb]); and sagittal and dorsal T1w-FLAIR postcontrast (7 to 10 minutes after injection). Images were acquired with a slice thickness ranging from 3 to 4 mm (0-mm interslice gap), a matrix of 320 × 224, and a field of view between 12 and 20 cm.

Image analysis—Five board-certified veterinary radiologists with variable experience and background in MRI interpretation (radiologist 1, MAD; radiologist 2, LB; radiologist 3, RD; radiologist 4, ARM; and radiologist 5, SLK) were asked to blindly score several imaging parameters for each patient using an online[e] database. Radiologist experience since board certification exceeded 10 years for 2 (radiologists 3 and 5), exceeded 5 years for 1 (radiologist 1), and was < 5 years for the remaining 2 (radiologists 2 and 4). Radiologists were from 2 continents, worked in 4 facilities, and had been trained at 4 universities.

The first task that was asked of the radiologists was to score the presence or absence of a notable brain lesion. If at least 1 notable lesion was detected, the distribution was scored as focal, multifocal, or diffuse. The lesion was then characterized (the largest if multiple lesions were present) for axial localization (intra-axial, extra-axial, or uncertain), mass effect (yes, no, or uncertain), cerebral edema (absent, mild, moderate, or marked), hemorrhagic component (present, absent, or uncertain), and cavitary component (present, absent, or uncertain). Brain contrast enhancement distribution was scored as absent, focal, multifocal, or diffuse. If more than 1 lesion was present, contrast enhancement characteristics of the largest were scored according to intensity (absent, mild, moderate, or marked), pattern (homogeneous, slightly heterogeneous, highly heterogeneous, or ring-like), definition of margins (highly ill-defined, slightly ill-defined, or well-defined), and presence of a dural tail (yes, no, or uncertain). Finally, readers selected their most likely diagnosis (no lesion, neoplastic, inflammatory, vascular, metabolic or toxic, or other [which included degenerative]) while grading their level of confidence: probable (50% to 75%), most likely (> 75% to 95%), or definite (> 95%). Once these parameters were all scored, patient information (ie, signalment and short clinical history) was automatically revealed to the readers in a new software window. Using these additional data, readers were asked to once more assign a final diagnosis and grade diagnostic confidence as previously described. Finally, readers selected the MRI sequence (regardless of imaging plane) that was most relevant in establishing their diagnosis.

Statistical analysis—A simple κ coefficient was calculated between each pair of readers for each of the following criteria: presence of a lesion; localization; mass effect; hemorrhagic component; presence of a cavitary area; distribution, pattern, and marginal definition of contrast enhancement; presence of a dural tail; most

SMALL ANIMALS/ AVIAN

determinant imaging sequence; and diagnosis with and without patient data provided.

Simple κ coefficients were also calculated between each pair of readers for the recorded pattern of enhancement (homogeneous, heterogeneous [slightly heterogeneous and highly heterogeneous combined], or ring-like) as well as recoded margins of enhancement (highly ill-defined, slightly ill-defined, or well-defined).

Weighted κ coefficients were calculated between each pair of readers for categories of ordinal variables, such as cerebral edema and degree of contrast enhancement. A simple κ coefficient was calculated for each reader's diagnosis by comparing his or her imaging diagnosis (obtained with and without patient data provided) with the final clinical diagnosis.

An adaptation of the κ evaluation grid by Landis and Koch[36] was used to qualify the level of inter-reader agreement **(Appendix)**.

Tests for symmetry were used to determine whether, for readers 3, 4, and 5 (ie, non–CHUM-UM radiologists not involved in the initial clinical reading of these patients), individually and combined, MRI categories of diagnoses were significantly ($P < 0.05$) different without and with patient data provided. Wilcoxon paired tests were used to determine whether the level of confidence in establishing a diagnosis, for each reader and when combining all 3 non–CHUV-UM readers, increased significantly when patient data were provided. A Cochran-Mantel-Haenszel test was performed to determine whether the level of diagnostic confidence significantly varied among diagnostic categories, established with and without patient data provided. Finally, for all readers as a group, an exact χ^2 test was obtained to determine whether a given MRI sequence was more likely to lead to a specific diagnosis. For all of these analyses, values of $P < 0.05$ were considered significant. All analyses were performed with standard statistical software.[f]

Diagnostic accuracy—For this analysis, only values obtained for the 3 readers from institutions other than the CHUV-UM were considered.

Results

Patient selection—Forty-four dogs met the inclusion criteria. Dogs had a median age of 6.2 years (range, 0.3 to 10.8 years) and median body weight of 23 kg (50.6 lb; range, 0.77 to 51.1 kg [1.7 to 112.4 lb]). Breeds represented included Akita (n = 1), Australian Shepherd (1), Basset Hound (1), Beagle (2), Bernese Mountain Dog (1), Border Collie (1), Boxer (4), Doberman Pinscher (1), English Springer Spaniel (1), German Shepherd Dog (1), Golden Retriever (4), Greyhound (1), Groenendael (1), Labrador Retriever (4), Maltese (2), Miniature Pinscher (1), Miniature Schnauzer (1), Pug (4), Siberian Husky (1), Shetland Sheepdog (1), Shih Tzu (3), Toy Poodle (1), Welsh Terrier (1), Yorkshire Terrier (2), and mixed (3). There were 25 males and 19 females. A histopathologic diagnosis was available in 33 of the 44 (75%) dogs and in the 18 (40.9%) dogs with neoplasia. An intracranial inflammatory process was identified in 18 (40.9%) dogs, which was confirmed on histologic examination in 9; 2 (4.6%) other dogs had a vascular event; and as confirmed on histologic examination, 3 (6.8%) dogs had no abnormalities found at necropsy, 2 (4.6%) dogs had a degenerative condition, and 1 (2.3%) dog had changes associated with a metabolic dysfunction. The relative number of animals retained in each category of diagnoses was kept unknown to radiologists to limit biases. Animals with nonfatal or medically controlled diseases were included to be representative of the caseload routinely imaged and to reduce selection bias toward animals with severe or surgically treatable conditions.

Interobserver agreement for lesion characteristics, contrast enhancement pattern, and diagnosis—κ Coefficient intervals were summarized **(Tables 1–3)**.

Diagnostic accuracy—Agreement between a reader's diagnoses and final clinical diagnoses was substantial, with ($0.65 < \kappa < 0.76$) or without ($0.65 < \kappa < 0.68$) patient data provided.

Influence of knowledge of patient data on MRI diagnosis—No significant ($P \geq 0.99$) difference was found between MRI diagnoses established without and with patient data for readers 3 to 5, all conditions considered. Also, their level of confidence did not change significantly ($P \geq 0.56$) whether patient data were provided or not. Without patient signalment, readers considered the final diagnosis to be probable 14.5% of the

Table 1—Interobserver agreement among 5 observers for lesion characteristics on retrospective brain MRI in 44 dogs on the basis of the range of κ statistics calculated for each pair of observers.

Variable	κ
Presence of lesion	> 0.64 to < 0.86
Localization	> 0.46 to < 0.80
Mass effect	> 0.51 to < 0.77
Hemorrhagic component	> 0.14 to < 0.53
Cavitary area	> 0.51 to < 0.72
Edema (weighted κ)	> 0.30 to < 0.52

Table 2—Interobserver agreement for contrast enhancement in the same dogs as in Table 1.

Variable	κ
Intensity (weighted κ)	> 0.41 to < 0.78
Distribution with each category compared separately	> 0.40 to < 0.70
Distribution with multifocal and diffuse categories combined	> 0.49 to < 0.72
Pattern with each category compared separately	> 0.16 to < 0.41
Pattern with slightly heterogeneous and frankly heterogeneous categories combined	> 0.28 to < 0.60
Margins with each category compared separately	> 0.22 to < 0.70
Margins with slightly ill-defined and frankly ill-defined categories combined	> 0.29 to < 0.70
Dural tail	> 0.22 to < 0.74

Table 3—Interobserver agreement for category of diagnosis in the same dogs as in Table 1.

Variable	κ
Diagnosis without patient information	> 0.56 to < 0.69*
Diagnosis with patient information	> 0.65 to < 0.69*
Most useful sequence	> 0.11 to < 0.71

*Values of κ are for 3 non–CHUV-UM radiologists.

Table 4—Frequency (%) of radiologists' choices for most useful imaging sequences for each diagnosis.

Variable	T1 precontrast	T2	T2-FLAIR	T1 postcontrast
Normal brain (n = 3)	0	21.4	35.7	42.9
Neoplastic (n = 18)	0	13.7	0	86.3
Inflammatory (n = 18)	0	11.0	41.1	48.0
Vascular (n = 2)	14.3	28.6	0	57.1
Metabolic or toxic (n = 1)	0	18.2	63.6	18.2
Other (n = 2)	11.1	44.4	0	44.4

Values may not sum to 100% because of rounding.

time, most likely 39.5% of the time, and definite 45.9% of the time. When patient signalment was provided, final diagnoses were considered probable 12.3% of the time, most likely 26.8% of the time, and definite 60.9% of the time.

Diagnostic confidence—When readers were considered individually, an association between the level of confidence and the type of diagnosis established without patient data was found for reader 4 ($P = 0.01$). For this reader, confidence scores were significantly increased when diagnosing neoplasia.

Most useful sequence—When all readers were considered as a group, a significant ($P < 0.001$) association was found between the most useful sequence and the category of diagnosis. The frequency of sequence selection for each individual category of diagnosis is provided **(Table 4)**. Radiologists overall selected the T1w postcontrast sequence as the most useful when identifying a normal brain (42.9%), neoplasia (86.3%), inflammation (48.0%), and a vascular event (57.1%), although there was not a consensus between the radiologists. On the other hand, the T2w-FLAIR sequence was considered the most useful when diagnosing metabolic disease or toxic insult (63.6%).

Discussion

κ Statistics have been extensively used to evaluate reading variability for diagnostic tests that rely on subjectivity,[37] which is inherent to most diagnostic imaging modalities. Whereas reading variability is sometimes reported in terms of frequency of agreement, the appropriate test also needs to take into account that observers will sometimes agree or disagree simply by chance[37] and this is precisely what κ coefficients determine. Simple κ coefficients are more appropriate for categorical nominal values (eg, yes or no), whereas weighted κ coefficients better reflect reading agreement for ordinal values (eg, mild, moderate, or severe).

The capacity of MRI to detect several canine brain diseases has been extensively described in the literature, contributing to make this modality the tool of choice for noninvasive assessment of brain disease in that species. While results of this diagnostic imaging modality (often through a radiologist's report) play a major role in prognosis and treatment planning, little is known about its accuracy. Moreover, investigations on interobserver variability, which have become more common in recent years, demonstrate that reading agreement can vary among tests in veterinary medicine, including MRI.[6,38–42] The present study highlights the fact that radiologists of different backgrounds may not agree on certain imaging features such as dural tail sign or the presence of edema, but that they overall agree satisfactorily on the category of brain disease when reviewing 1.5-T MRI images of canine brains obtained with a standard acquisition protocol.

In this study, brain MRI images of 44 dogs were reviewed, of which 33 had a histopathologic diagnosis. Eleven patients (9 inflammatory and 2 vascular events) were entered in the study without histologic confirmation. To confidently establish a diagnosis in these dogs, other tests were considered: CSF analysis as well as follow-up neurologic examination and MRI. Cerebrospinal fluid analysis is known to be the most useful test to identify intracranial inflammation[2,43] and was used to support the diagnosis in our patients. On the other hand, because neoplasia can lead to an inflammatory CSF,[2,44] repeated neurologic evaluations were made of all dogs with clinically suspected inflammatory diseases, and MRI was repeated in 4 of these patients. Follow-up MRI showed either complete resolution of initial lesions 6 weeks (n = 1 patient) to 6 months (2) after initiation of medical treatment or showed considerable improvement of lesions 1 year after initial MRI (1). Both dogs with clinically suspected brain infarction (n = 2) were categorized on the basis of a combination of results of diffusion sequences performed at the time of initial MRI examination and follow-up MRI images obtained 4 (n = 1) and 7 weeks (1) later. These 2 dogs did not receive any treatment after initial evaluation and had recovered by the time of follow-up examination. The addition of these patients appeared essential to the authors to ensure a better representation of a typical clinical caseload. Most animals with treatable or self-limiting diseases undergo MRI without confirmation of their diagnosis at the time of histologic examination. Removing these patients from our study would have resulted in reading biases.

Sequences and planes selected in this study reflect the protocol routinely used at the CHUV-UM. As protocols might vary among imaging centers, it is possible that selected sequences, such as T1w-FLAIR, induced a certain bias, lowering confidence of some of our external readers not familiar with such sequences. Also, other MRI sequences such as T2*-GRE were often obtained at the time of initial MRI evaluation, but only 4 sequences in different planes (for a total of 8 sequences) were made available for readers in this study. T2-weighted–FSE and T1w-FLAIR pre- and postcontrast sequences were selected for their recognized importance[45] in brain imaging and their common use in veterinary medicine.[2,22,30,46,47] Whereas postcontrast T1w images were acquired in 3 planes, precontrast T1w-FLAIR images were only acquired in transverse planes, as recommended.[48] A T2w-FLAIR sequence was added for its proven sensitivity in identifying inflammatory lesions[4,41] and for its overall benefit in distinguishing tissue T2 hyperintensity from CSF. It was also chosen to keep the imaging protocol constant among patients to reduce reading biases. Indeed, the addition of T2*-GRE and diffusion sequences in a subset of patients may have enabled readers to better discriminate between hemorrhagic or ischemic cerebrovascular accidents, respectively, because these sequences are most often used for these purposes.[25,26] Hence, we acknowledge the fact that reading agreement in recognizing hemorrhage may have been superior if these specific sequences had been made available to readers.

Radiologists were selected for their diverse experience in MRI reading as well as for their different backgrounds. Of the 5 readers, 2 were from the CHUV-UM (readers 1 and 2) and could have been previously involved in the interpretation of some of the cases. To reduce bias, both of these readers were excluded from the assessment of diagnostic accuracy as well as for the evaluation of diagnostic confidence and influence of patient data, since these criteria are linked to the diagnosis. In fact, although the readers were blinded to the images, they might have recognized specific cases, which could have influenced their final diagnosis. However, we decided their scores for lesion characteristics reflected their potential familiarity with these cases and did not substantially influence their interpretation of specific imaging criteria. Furthermore, their selection of the most useful sequence for each diagnosis was retained.

Substantial to almost perfect agreement was obtained when readers were asked to identify a lesion. Describing lesion distribution, whether focal, multifocal, or diffuse, is an important part of its classification and helps establish a final diagnosis. Definitions for these types of distribution were not provided to the readers in this study because the authors' interest was to evaluate whether radiologists use these terms consistently. The strong interobserver agreement in this category ($0.64 < \kappa < 0.86$) suggests that radiologists seem to classify lesion distribution consistently.

Localization of a lesion (ie, extra- vs intra-axial) is important in identifying a specific diagnosis.[49–51] Different tumors, for instance, are found within the brain parenchyma (intra-axial) while others will affect mostly extra-axial structures. Agreement was moderate to almost perfect when readers were asked to specify whether a lesion was intra-axial, extra-axial, or of uncertain localization. Nevertheless, this agreement varied considerably ($0.46 < \kappa < 0.80$) among pairs of readers. This variation could hardly be explained by the difference in experience because the lowest κ score was obtained between the 2 most experienced readers. This discrepancy may be explained, at least in part, by the tendency of some readers to score as uncertain for equivocal patients, whereas others prefer to commit themselves. Hence, this additional category of localization (ie, uncertain) might have contributed to artificially reducing the reading agreement. Another explanation could be that, in some patients, the lesion had both intra- and extra-axial components, making the choice difficult for readers. In fact, a radiologist might classify the lesion on the basis of its assumed origin while another might consider the location of the largest part of the lesion when classifying it.

Characterization of a mass effect was also evaluated in this study, as this criterion can sometimes help distinguish different diseases.[8,52] Generally speaking, a mass effect is present when there is shift of the brain parenchyma or when compression of the ventricular system is observed.[8,30] Readers of this study were not provided with any specific guideline for this criterion, and their scoring agreement was only moderate to substantial. Additionally, some readers had a tendency to score this criterion as uncertain more often than did others. These results suggest a lack of consensus on the definition of the presence and magnitude of a mass effect on brain MRI images in dogs.

The presence of hemorrhage within the main lesion was recognized with low and quite variable agreement. This is explained, at least in part, by the fact that only basic sequences were made available for the readers. It is recognized that MRI signal intensity of hemorrhage on T1w- and T2w-FSE images varies according to its time of onset relative to imaging and that it can be difficult to identify on these conventional sequences, especially in the acute phase.[21–26] Gradient echo sequences and especially T2*-GRE are, on the other hand, more sensitive, and their results do not depend on the hemorrhage time of onset for at least the first 8 days following the incident.[22,25,26,53] Hence, the low agreement between radiologists asked to recognize hemorrhage further supports the fact that specific sequences should be used when such a component is suspected and particularly when its presence influences the diagnosis.

Interestingly, the presence of a cavitary area in the primary brain lesion was more consistently recognized by radiologists ($0.51 < \kappa < 0.72$). This was surprising when considering the lack of a specific definition for this characteristic and the difficulty in recognizing a cavitary area when the lesion's contours are not well-defined. On the other hand, the fact that this component had low prevalence among readers may have positively influenced the agreement.

Evaluation of cerebral edema led to fair to moderate ($0.30 < \kappa < 0.52$) agreement among radiologists. Grading more than recognizing cerebral edema was initially thought to be the cause of low agreement because grading scores (ie, mild, moderate. and marked) may be interpreted differently. However, even with weighted κ, which take into account the gap between ordinal scores (eg, mild and moderate would result in better κ value than would mild and severe), the agreement remained low. This reading variability is probably explained by the fact that edema is not always straightforwardly distinguished from neoplastic and inflammatory infiltrates or from necrosis, which also produces ill-defined hyperintensity on T2w-FSE and T2w-FLAIR sequences. As edema should not increase following contrast enhancement,[54,55] evaluation of postcontrast sequences is important to better differentiate this process from lesions associated with breakage of the blood-brain barrier or neovascularisation. Our results justify the need for establishing clearer reading guidelines for this important imaging criterion.

Numerous contrast enhancement characteristics were evaluated in this study. However, as these characteristics were only scored when readers recognized the presence of enhancement, κ values could be calculated only for a subset of patients, which varied for the various parameters assessed. That being said, readers moderately to substantially agreed on the presence and intensity of enhancement and, as was the case before contrast enhancement, readers did not have difficulty agreeing on multifocal and diffuse enhancement, suggesting both categories are distinct for most readers. Conversely, the pattern of enhancement appeared more difficult for readers to agree on. A reason for this modest agreement ($0.16 < \kappa < 0.41$) might be that the heterogeneous category was divided into subcategories: slightly and highly heterogeneous. However, even when these subcategories were combined, interobserver agreement remained below the moderate range. Definition of contour of contrast enhancement was also not consistently categorized by radiologists. κ Values varied greatly ($0.22 < \kappa < 0.70$), even when both ill-defined categories (ie, slightly and highly) were combined ($0.29 < \kappa < 0.70$), indicating that the terms well-defined and ill-defined were not consistently used.

Although most typically associated with meningiomas, the presence of a dural tail (ie, linear meningeal thickening and enhancement adjacent to and in continuity with a peripherally located cranial mass[56]) has been described with other diseases including inflammatory conditions and other neoplasms.[29,57] In fact, the dural tail sign is no longer considered specific to a particular pathological process.[29,57] Although very few studies have looked at this parameter in veterinary medicine, most concluded that this sign is strongly predictive of meningiomas[14,29] but difficult for veterinary radiologists to appreciate.[29,58] This last observation was supported by our study in which interobserver agreement on the presence of a dural tail was quite variable ($0.22 < \kappa < 0.74$). Interestingly, on 4 occasions in this study, a reader chose an inflammatory diagnosis while recognizing a dural tail sign.

Once all characteristics were looked at individually, readers were asked to provide a diagnosis. As previously stated, 2 of the radiologists (readers 1 and 2) were excluded from the portion of the study that evaluated diagnostic accuracy because they might have been involved in the initial reading of some of the 44 patients. Moderate to substantial agreement was found among pairs of the remaining readers for establishing a diagnosis. Moreover, the agreement did not notably vary when patient information was provided to these readers. Interestingly, readers were more consistent when choosing a category of diagnosis as opposed to scoring the most specific imaging features. This suggests that, at least in this study, even if the most specific MRI features of diseases are variably interpreted, radiologists most often agree on a final category of diagnosis (ie, normal, neoplastic, inflammatory, or vascular condition).

Although MRI examinations are interpreted as an ensemble, some sequences might have a greater impact on the diagnosis of specific types of diseases or even globally. In this study, the selection of the most useful sequence for each imaging patient was not consistent among the 5 readers ($0.11 < \kappa < 0.71$). This discrepancy may be related to the fact that these radiologists came from 4 imaging centers and might have had different experience with some of the sequences included in this study. Nevertheless, a significant ($P < 0.001$) association between the most determinitive sequence and the established diagnosis was recognized. For instance, 86.3% of the time, when a lesion was characterized as neoplastic, the reader chose the T1w postcontrast sequence as being the most useful sequence, whereas the T2w-FSE sequence was selected 13.7% of the time. Readers might have relied on the fact that a larger proportion of neoplastic versus nonneoplastic conditions improve after gadolinium injection.[8,14] On the other hand, when a diagnosis of inflammatory disease was made, readers were divided in their selection of the most useful sequence (ie, mainly between the T1w postcontrast [48.0%] and T2w-FLAIR [41.1%]) sequences.

Studies in veterinary medicine showed higher sensitivity of FLAIR (ie, T2w-FLAIR) sequences for the detection of occult inflammatory lesions that were not visible on T2w or T1w postcontrast sequences.[4,41] This could explain why, in the present study, the T2w-FLAIR sequence was almost equally selected as the most useful sequence for making a diagnosis of inflammation, when compared with contrast-enhanced T1w-FLAIR sequence.

A symmetry test revealed that diagnoses did not significantly change when patient data were provided for any of the radiologists and that their level of confidence did not increase with that data provided. This finding suggests that patient information does not play an important role in the establishment of the MRI diagnosis, at least when this diagnosis is applied to a category of diseases. On the other hand, knowledge of patient data may have greater impact on the radiologist's opinion when, for instance, a neoplasia subtype needs to be determined (eg, meningioma vs glioma). Owing to the limited number of subtypes and great variety in each category of diseases, it was not possible to further study the agreement of readers for these more specific diagnoses.

This study represents the first published data on interobserver reading agreement for canine brain MRI diagnosis and imaging characteristics. It highlights the fact that radiologists of different backgrounds may not agree on certain imaging features such as dural tail sign, edema, and pattern of enhancement, justifying the need for standardized definitions and scoring schemes for each of these imaging characteristics. Furthermore, this study also shows that radiologists do not agree on recognition of hemorrhage on conventional sequences and that specific sequences such as T2w*-GRE sequences should be added to better identify this characteristic. Nevertheless, this study confirms that radiologists are reasonably consistent and accurate for choosing a category of diagnosis when interpreting canine brain MRI images acquired with a standard 1.5-T MRI protocol.

a. Signa HDxt, General Electric Medical System, Mississauga, ON, Canada.
b. 8 CH Knee Coil, Invivo Corp, Gainesville, Fla.
c. 8 CH CTL Spine Coil, USA Instruments Inc, Aurora, Ohio.
d. Multihance, Bracco Diagnostics Inc, Vaughan, ON, Canada.
e. Office Access, Office 2007, Microsoft Canada, Mississauga, ON, Canada.
f. SAS, version 9.2, SAS Institute Inc, Cary, NC.

References

1. Cherubini G, Platt S, Anderson T, et al. Characteristics of magnetic resonance images of granulomatous meningoencephalomyelitis in 11 dogs. *Vet Rec* 2006;159:110–115.

2. Lamb C, Croson P, Cappello R, et al. Magnetic resonance imaging findings in 25 dogs with inflammatory cerebrospinal fluid. *Vet Radiol Ultrasound* 2005;46:17–22.
3. Von Praun F, Matiasek K, Grevel V, et al. Magnetic resonance imaging and pathologic findings associated with necrotizing encephalitis in two Yorkshire Terriers. *Vet Radiol Ultrasound* 2006;47:260–264.
4. Cherubini G, Platt S, Howson S, et al. Comparison of magnetic resonance imaging sequences in dogs with multi-focal intracranial disease. *J Small Anim Pract* 2008;49:634–640.
5. Higginbotham MJ, Kent M, Glass EN. Noninfectious inflammatory central nervous system diseases in dogs. *Compend Contin Educ Vet* 2007;29:488–497, 501.
6. Young BD, Levine JM, Fosgate GT, et al. Magnetic resonance imaging characteristics of necrotizing meningoencephalitis in Pug dogs. *J Vet Intern Med* 2009;23:527–535.
7. Granger N, Smith PM, Jeffery ND. Clinical findings and treatment of non-infectious meningoencephalomyelitis in dogs: a systematic review of 457 published cases from 1962 to 2008. *Vet J* 2010;184:290–297.
8. Thomas WB, Wheeler SJ, Kramer R, et al. Magnetic resonance imaging features of primary brain tumors in dogs. *Vet Radiol Ultrasound* 1996;37:20–27.
9. Hecht S, Adams WH. MRI of brain disease in veterinary patients part 2: acquired brain disorders. *Vet Clin North Am Small Anim Pract* 2010;40:39–63.
10. Rodenas S, Pumarola M, Gaitero L, et al. Magnetic resonance imaging findings in 40 dogs with histologically confirmed intracranial tumours. *Vet J* 2011;187:85–91.
11. Westworth DR, Dickinson PJ, Vernau W, et al. Choroid plexus tumors in 56 dogs (1985–2007). *J Vet Intern Med* 2008;22:1157–1165.
12. Sturges BK, Dickinson PJ, Bollen AW, et al. Magnetic resonance imaging and histological classification of intracranial meningiomas in 112 dogs. *J Vet Intern Med* 2008;22:586–595.
13. Snyder JM, Shofer FS, Van Winkle TJ, et al. Canine intracranial primary neoplasia: 173 cases (1986–2003). *J Vet Intern Med* 2006;20:669–675.
14. Cherubini GB, Mantis P, Martinez TA, et al. Utility of magnetic resonance imaging for distinguishing neoplastic from non-neoplastic brain lesions in dogs and cats. *Vet Radiol Ultrasound* 2005;46:384–387.
15. Lipsitz D, Higgins RJ, Kortz GD, et al. Glioblastoma multiforme: clinical findings, magnetic resonance imaging, and pathology in five dogs. *Vet Pathol* 2003;40:659–669.
16. Bagley RS, Wheeler SJ, Klopp L, et al. Clinical features of trigeminal nerve-sheath tumor in 10 dogs. *J Am Anim Hosp Assoc* 1998;34:19–25.
17. Kraft SL, Gavin PR, DeHaan C, et al. Retrospective review of 50 canine intracranial tumors evaluated by magnetic resonance imaging. *J Vet Intern Med* 1997;11:218–225.
18. Duesberg CA, Feldman EC, Nelson RW, et al. Magnetic resonance imaging for diagnosis of pituitary macrotumors in dogs. *J Am Vet Med Assoc* 1995;206:657–662.
19. Bagley RS, Kornegay JN, Lane SB, et al. Cystic meningiomas in 2 dogs. *J Vet Intern Med* 1996;10:72–75.
20. Goncalves R, Carrera I, Garosi L, et al. Clinical and topographic magnetic resonance imaging characteristics of suspected thalamic infarcts in 16 dogs. *Vet J* 2011;188:39–43.
21. Garosi L, McConnell J, Platt S, et al. Results of diagnostic investigations and long-term outcome of 33 dogs with brain infarction (2000–2004). *J Vet Intern Med* 2005;19:725–731.
22. Garosi L, McConnell J, Platt S, et al. Clinical and topographic magnetic resonance characteristics of suspected brain infarction in 40 dogs. *J Vet Intern Med* 2006;20:311–321.
23. Berg, JM, Joseph, RJ. Cerebellar infarcts in two dogs diagnosed with magnetic resonance imaging. *J Am Anim Hosp Assoc* 2003;39:203–207.
24. Garosi LS, McConnell JF. Ischaemic stroke in dogs and humans: a comparative review. *J Small Anim Pract* 2005;46:521–529.
25. McConnell JF, Garosi L, Platt SR. Magnetic resonance imaging findings of presumed cerebellar cerebrovascular accident in twelve dogs. *Vet Radiol Ultrasound* 2005;46:1–10.
26. Wessmann A, Chandler K, Garosi L. Ischaemic and haemorrhagic stroke in the dog. *Vet J* 2009;180:290–303.
27. Bathen-Noethen A, Stein VM, Puff C, et al. Magnetic resonance imaging findings in acute canine distemper virus infection. *J Small Anim Pract* 2008;49:460–467.
28. Garosi L, Dawson A, Couturier J, et al. Necrotizing cerebellitis and cerebellar atrophy caused by *Neospora caninum* infection: magnetic resonance imaging and clinicopathologic findings in seven dogs. *J Vet Intern Med* 2010;24:571–578.
29. Graham JP, Newell SM, Voges AK, et al. The dural tail sign in the diagnosis of meningiomas. *Vet Radiol Ultrasound* 1998;39:297–302.
30. Hecht S, Adams WH. MRI of brain disease in veterinary patients part 1: basic principles and congenital brain disorders. *Vet Clin North Am Small Anim Pract* 2010;40:21–38.
31. De Haan CE, Kraft SL, Gavin PR, et al. Normal variation in size of the lateral ventricles of the Labrador Retriever dog as assessed by magnetic resonance imaging. *Vet Radiol Ultrasound* 1994;35:83–86.
32. Kii S, Uzuka Y, Taura Y, et al. Magnetic resonance imaging of the lateral ventricles in Beagle-type dogs. *Vet Radiol Ultrasound* 1997;38:430–433.
33. Vite CH, Insko EK, Schotland HM, et al. Quantification of cerebral ventricular volume in English Bulldogs. *Vet Radiol Ultrasound* 1997;38:437–443.
34. Esteve-Ratsch B, Kneissl S, Gabler C. Comparative evaluation of the ventricles in the Yorkshire Terrier and the German Shepherd Dog using low-field MRI. *Vet Radiol Ultrasound* 2001;42:410–413.
35. Thomas WB. Evaluation of veterinary patients with brain disease. *Vet Clin North Am Small Anim Pract* 2010;40:1–19.
36. Landis J, Koch G. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics* 1977;33:363–374.
37. Viera A, Garrett J. Understanding interobserver agreement: the kappa statistic. *Fam Med* 2005;37:360–363.
38. Barrett E, Barr F, Owen M, et al. A retrospective study of the MRI findings in 18 dogs with stifle injuries. *J Small Anim Pract* 2009;50:448–455.
39. Naude SH, Lambrechts NE, Wagner WM, et al. Association of preoperative magnetic resonance imaging findings with surgical features in Dachshunds with thoracolumbar intervertebral disk extrusion. *J Am Vet Med Assoc* 2008;232:702–708.
40. Suwankong N, Voorhout G, Hazewinkel HA, et al. Agreement between computed tomography, magnetic resonance imaging, and surgical findings in dogs with degenerative lumbosacral stenosis. *J Am Vet Med Assoc* 2006;229:1924–1929.
41. Benigni L, Lamb C. Comparison of fluid-attenuated inversion recovery and T2-weighted magnetic resonance images in dogs and cats with suspected brain disease. *Vet Radiol Ultrasound* 2005;46:287–292.
42. Schulz KS, Walker M, Moon M, et al. Correlation of clinical, radiographic, and surgical localization of intervertebral disc extrusion in small-breed dogs: a prospective study of 50 cases. *Vet Surg* 1998;27:105–111.
43. Tipold A. Diagnosis of inflammatory and infectious diseases of the central nervous system in dogs: a retrospective study. *J Vet Intern Med* 1995;9:304–314.
44. Bohn AA, Wills TB, West CL, et al. Cerebrospinal fluid analysis and magnetic resonance imaging in the diagnosis of neurologic disease in dogs: a retrospective study. *Vet Clin Pathol* 2006;35:315–320.
45. Al-Saeed O, Ismail M, Athyal RP, et al. T1-weighted fluid-attenuated inversion recovery and T1-weighted fast spin-echo contrast-enhanced imaging: a comparison in 20 patients with brain lesions. *J Med Imaging Radiat Oncol* 2009;53:366–372.
46. Tucker RL, Gavin PR. Brain imaging. *Vet Clin North Am Small Anim Pract* 1996;26:735–758.
47. Thomson CE, Kornegay JN, Burn RA, et al. Magnetic resonance imaging—a general overview of principles and examples in veterinary neurodiagnosis. *Vet Radiol Ultrasound* 1993;34:2–17.
48. Robertson, I. Optimal magnetic resonance imaging of the brain. *Vet Radiol Ultrasound* 2011;52(suppl 1):S15–S22.
49. Kraft SL, Gavin PR. Intracranial neoplasia. *Clin Tech Small Anim Pract* 1999;14:112–123.
50. Robertson ID. Magnetic resonance imaging features of brain

disease in small animals. In: Thrall DE, ed. *Textbook of veterinary diagnostic radiology*. 5th ed. St Louis: Saunders Elsevier, 2007;142–159.
51. Thomas WB. Nonneoplastic disorders of the brain. *Clin Tech Small Anim Pract* 1999;14:125–147.
52. Cervera V, Mai W, Vite CH, et al. Comparative magnetic resonance imaging findings between gliomas and presumed cerebrovascular accidents in dogs. *Vet Radiol Ultrasound* 2011;52:33–40.
53. Siddiqui FM, Bekker SV, Qureshi AI. Neuroimaging of hemorrhage and vascular defects. *Neurotherapeutics* 2011;8:28–38.
54. Tidwell AS. Principles of computed tomography and magnetic resonance imaging. In: Thrall DE, ed. *Textbook of veterinary diagnostic radiology*. 5th ed. St Louis: Saunders Elsevier, 2007;50–77.
55. Tidwell AS, Jones JC. Advanced imaging concepts: a pictorial glossary of CT and MRI technology. *Clin Tech Small Anim Pract* 1999;14:65–111.
56. Rokni-Yazdi H, Sotoudeh H. Prevalence of “dural tail sign” in patients with different intracranial pathologies. *Eur J Radiol* 2006;60:42–45.
57. Guermazi A, Lafitte F, Miaux Y, et al. The dural tail sign—beyond meningioma. *Clin Radiol* 2005;60:171–188.
58. Hathcock JT. Low field magnetic resonance imaging characteristics of cranial vault meningiomas in 13 dogs. *Vet Radiol Ultrasound* 1996;37:257–263.

Appendix

Scale used to qualify the level of interobserver agreement, determined on the basis of the κ statistic, for retrospective evaluation of brain MRI images from 44 dogs.

κ	**Agreement**
> 0.80	Almost perfect
> 0.60 to 0.80	Substantial
> 0.40 to 0.60	Moderate
> 0.20 to 0.40	Fair
> 0.00 to 0.20	Slight
0	Poor