

Standard Article

J Vet Intern Med 2017;31:778–783

Interobserver Agreement Using Histological Scoring of the Canine Liver

J.A. Lidbury , A. Rodrigues Hoffmann, R. Ivanek, J.M. Cullen, B.F. Porter, F. Oliveira, T.J. Van Winkle, G.C. Grinwis, J.S. Sucholdolski, and J.M. Steiner

Background: Grading schemes for the assessment of hepatic fibrosis and necroinflammatory activity in humans previously have been applied to dogs with chronic hepatitis. Interobserver agreement is a desirable characteristic for any histological scoring scheme.

Hypothesis/Objectives: To assess interobserver agreement associated with pathologists using a previously published histological scoring scheme to assess hepatic fibrosis and necroinflammatory activity in dogs and to compare fibrosis scores assigned to serial sections stained with hematoxylin & eosin (H&E) and picrosirius red.

Animals: Histological sections of liver from 50 dogs with variable degrees of hepatic fibrosis and necroinflammatory activity were selected from institutional tissue archives.

Methods: Six board-certified veterinary anatomic pathologists assigned fibrosis and necroinflammatory activity scores to the histological sections. The multiuser kappa statistic was calculated to assess interobserver agreement. Fibrosis stage assigned to serial sections stained with picrosirius red and H&E was compared using the Wilcoxon signed-rank test.

Results: Multiuser kappa statistics for assessment of fibrosis and necroinflammatory activity from H&E-stained sections were 0.35 and 0.16, respectively. There was no difference in median fibrosis scores assigned to serial section stained with H&E and picrosirius red ($P = .248$).

Conclusions and Clinical Importance: There was fair interobserver agreement when pathologists assessed fibrosis and poor agreement when they assessed necroinflammatory activity. This suboptimal agreement must be taken into account by clinicians making decisions based on histology reports of the liver and in the design of studies evaluating these findings. To decrease this variability, ideally >1 pathologist should evaluate each section.

Key words: Chronic hepatitis; Dog; Hepatic fibrosis; histological scoring.

Several histological grading systems have been developed for assessment of hepatic necroinflammatory activity and fibrosis in human patients with chronic hepatitis, including the Knodell,¹ Ishak,² and

From the Gastrointestinal Laboratory, (Lidbury, Sucholdolski, Steiner); Department of Veterinary Pathobiology, (Rodrigues Hoffmann, Porter); Department of Veterinary Integrative Biosciences, College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, TX (Ivanek); North Carolina College of Veterinary Medicine, North Carolina State University, Raleigh, NC (Cullen); Antech Diagnostics, College Station, TX (Oliveira); Department of Pathobiology, School of Veterinary Medicine, University of Pennsylvania, Philadelphia, PA (Van Winkle); and the Department of Pathobiology, Faculty of Veterinary Medicine, University of Utrecht, Utrecht, The Netherlands (Grinwis).

Dr. Ivanek's Present address is Department of Population Medicine & Diagnostic Services, College of Veterinary Medicine, Cornell University, Ithaca, NY.

Where the work was done: The study was conducted at Texas A&M University.

Meeting at which the paper was presented: A part of the study was being presented at the 2016 ACVIM Forum, Denver, Colorado.

Corresponding author: J. Lidbury, Gastrointestinal Laboratory, College of Veterinary Medicine & Biomedical Sciences, Texas A&M University, 4474 TAMU, College Station, TX 77843; e-mail: jlidbury@cvm.tamu.edu.

Submitted November 3, 2016; Revised January 16, 2017; Accepted January 30, 2017.

Copyright © 2017 The Authors. Journal of Veterinary Internal Medicine published by Wiley Periodicals, Inc. on behalf of the American College of Veterinary Internal Medicine.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

DOI: 10.1111/jvim.14684

Abbreviations:

CC	central–central
H&E	hematoxylin and eosin
NA	not applicable
PC	portal–central
PP	portal–portal
κ'	Cohen's weighted kappa statistic
κ	Cohen's kappa statistic

METAVIR³ schemes. Necroinflammatory activity grade encompasses various patterns of hepatocellular necrosis and inflammation, providing information regarding activity of the disease process, whereas the fibrosis stage gives an indication of disease chronicity.⁴ Despite use of histological scoring schemes in human patients with liver disease, currently there is no widely accepted scheme for use in dogs. A semiquantitative scoring system encompassing necroinflammatory changes, apoptosis, and fibrosis previously was developed for use in dogs with primary hepatitis.⁵ More recently, several studies in dogs with chronic hepatitis^{6,7} have used a histological scoring scheme for necroinflammatory grade and fibrosis stage that was adapted from the Ishak scheme.⁸

It has been proposed that histological scoring schemes fulfill several criteria, including interobserver agreement, intra-observer agreement, and clinical relevance.^{2,4} Suboptimal agreement among pathologists evaluating histological sections prepared from intestinal biopsy specimens from dogs has previously been documented,^{9,10} and 1 study found a lack of interobserver agreement in the morphologic diagnosis assigned

to needle and wedge liver biopsy specimens from dogs.¹¹ Other studies involving a single pathologist suggest that the histopathologic interpretation of a canine liver sample is unlikely to vary if the specimen contains at least 3–12 portal triads, regardless of the biopsy technique,¹² and that the likelihood of obtaining a representative sample is increased when multiple liver lobes are sampled.¹³ To our knowledge, interobserver agreement associated with the histological scoring of fibrosis and necroinflammatory activity from hepatic biopsy specimens collected from dogs has not been reported previously.

Cohen's kappa statistic (κ) frequently is used to estimate agreement of observers for data on nominal scales.¹⁴ A κ of zero represents no agreement beyond that due to chance, and a κ of 1.0 represents complete agreement. For nominal scoring systems, partial agreement between users may be taken into account.¹⁵ A weighted kappa statistic (κ') accounts for partial agreement by assigning weights to different levels of disagreement.¹⁴

The primary objective of our study was to assess interobserver agreement associated with the use of a scoring scheme to evaluate hepatic fibrosis in dogs. The secondary objectives were to assess interobserver agreement associated with the use of a scoring scheme to evaluate hepatic necroinflammatory activity and to compare fibrosis scores assigned to serial sections of the same biopsy specimen stained with hematoxylin and eosin (H&E) and picrosirius red. We hypothesized that pathologists assign higher fibrosis scores to serial sections of canine liver stained with picrosirius red to those stained with H&E.

Materials and Methods

Case Material

Fifty paraffin-embedded specimens of canine liver with variable degrees of fibrosis and necroinflammatory activity were selected from tissue archives. The sections were selected primarily to represent the full severity range of hepatic fibrosis by a clinician (JAL) who read the preexisting histopathology reports for these cases. This was done to avoid having poorly represented or unrepresented lesion severities, which could lead to falsely lower kappa values or make it impossible assess agreement for lesions of those severities, respectively. Selected case material has been used in other agreement studies.^{16,17} Overall, 36 dogs had chronic hepatitis, 11 were considered to be free from liver disease, and 3 had hepatic changes associated with congenial portosystemic shunts. No dogs were euthanized or underwent liver biopsy for study related reasons. Seventeen wedge biopsy specimens were collected at necropsy, 17 were collected during laparotomy, 11 were collected during laparoscopy with forceps (typically 4–6), and 5 were collected during laparotomy using skin biopsy punches. We opted to use wedge biopsy samples collected at necropsy, laparoscopically collected samples, and those collected during laparotomy so that interobserver agreement could be assessed when evaluating specimens of adequate size. Two serial 4–5 μm sections of liver were cut from the paraffin-embedded tissue and mounted onto separate microscope slides. One section was stained manually with picrosirius red^a and counter-stained with Weigert's iron hematoxylin.^b The sections were stained in as few batches as possible (3) to maintain consistency. The other

section was stained routinely with H&E using an automated slide stainer.

Histological Assessment

A number from 1 to 100 was randomly assigned to each section (50 stained with H&E and 50 stained with picrosirius red). The sections then were relabeled with this number as their only identifier. A digital image of each whole section was captured with a slide-scanning microscope^c using the 40 \times objective lens. Six board-certified veterinary anatomic pathologists evaluated the images of the scanned sections (3) or the slides (3), depending on their location. At the time, the study was performed 3 of the pathologists had ≥ 20 years of post-residency experience, 1 had 12 years of post-residency experience, and 2 had 5 years of post-residency experience. One of the pathologists helped to develop the scoring scheme. The remaining 5 pathologists were unfamiliar with the scheme and did not receive specific training before this study. For the sections of liver stained with H&E, the pathologists evaluated the stage of fibrosis and grade of necroinflammatory activity using a histological scoring scheme that was adapted from the human Ishak scheme.⁸ For the sections stained with picrosirius red, the pathologists scored stage of fibrosis only. According to this scoring scheme, necroinflammatory activity is graded as 0: absent, 1: slight, 2: mild, 3: moderate, 4: marked, or 5: very marked, and fibrosis is graded 0: absent, 1: mild, 2: moderate, 3: marked, or 4: very marked (Appendix 1). During the scoring process, the pathologists were not aware of the identity of the sections they were assessing, the scores assigned by the other pathologists, or the results of image analysis.

Statistical Analysis

To evaluate interobserver agreement, κ for each pair of observers was calculated. Kappa and κ' for multiple observers was calculated for each scoring category. Kappa was weighted as follows: 1 – (difference in rating between 2 raters/(maximum number of possible ratings – 1)). Kappa and κ' values were interpreted using the following guidelines: poor agreement <0.20, fair agreement 0.21–0.40, moderate agreement 0.41–0.60, good agreement 0.61–0.80, and very good agreement 0.81–1.00.¹⁴ Interobserver agreement also was summarized by calculating the frequency of scores assigned by each of the 15 possible pairs of pathologists. This analysis was performed using a commercially available software package.^d The median fibrosis and necroinflammatory scores assigned by the 6 pathologists for each section were compared using Friedman's test, followed by Dunn's post-test as appropriate. The median fibrosis stage assigned to each case for the sections stained with picrosirius red and H&E was compared using the Wilcoxon signed-rank test. This analysis was performed using another software package.^e Statistical significance level was set at $\alpha < 0.05$.

Results

All 100 sections were deemed to be of adequate to be size and quality for analysis by the 6 veterinary pathologists.

Fibrosis (Stage)

Agreement between the pairs of pathologists assigning scores for hepatic fibrosis to sections stained with H&E is presented in Table 1. The median (minimum–maximum) κ for each pair was 0.41 (0.14–0.56).

Table 1. Pairwise comparison of kappa statistics for the assessment of fibrosis from hematoxylin and eosin-stained sections.

Observer	1	2	3	4	5	6
1	NA	0.47	0.48	0.28	0.41	0.41
2	0.47	NA	0.56	0.17	0.42	0.35
3	0.48	0.56	NA	0.14	0.42	0.48
4	0.28	0.17	0.14	NA	0.18	0.17
5	0.41	0.42	0.42	0.18	NA	0.44
6	0.41	0.35	0.48	0.17	0.44	NA

NA, not applicable.

Multiusers κ (95% confidence interval [CI]) was 0.35 (0.26–0.44), and multiusers κ' was 0.59 (0.50–0.70). Assignment of fibrosis scores by the 15 possible pairings of pathologists to sections stained with H&E is presented in Table 2. The pairs of pathologists were in complete agreement 48.8% of the time, differed by 1 score level 40.5% of the time, and differed by 1 score level 10.7% of the time. There was a significant difference in fibrosis scores assigned to H&E-stained sections by the 6 pathologists ($P < .0001$), and significant differences were found between scores for 5 of 15 possible pathologist pairings.

Agreement between the pairs of pathologists assigning scores for hepatic fibrosis to sections stained with picrosirius red is presented in Table 3. The median (minimum–maximum) κ for each pair was 0.40 (0.22–0.56).

Table 2. Assignment of fibrosis scores by pathologist pairings for hematoxylin and eosin-stained sections.

Score	Absent	Mild	Moderate	Marked	Very Marked
Absent	9.9%	NA	NA	NA	NA
Mild	12.9%	7.2%	NA	NA	NA
Moderate	5.3%	9.7%	14.7%	NA	NA
Marked	0.7%	2.4%	11.9%	3.2%	NA
Very marked	0.0%	0.5%	1.7%	6.0%	13.9%

NA, not applicable.

The percentages represent the frequency at which the 15 possible pathologist pairings assigned hepatic fibrosis scores to the sections. The pairs of pathologists were in complete agreement 48.8% of the time (dark gray), differed by 1 score level 40.5% of the time (light gray), and differed by >1 score level 10.7% of the time (white).

Table 3. Pairwise comparison of kappa statistics for the assessment of fibrosis from picrosirius red-stained sections.

Observer	1	2	3	4	5	6
1	NA	0.50	0.40	0.24	0.29	0.48
2	0.50	NA	0.45	0.29	0.43	0.40
3	0.40	0.45	NA	0.49	0.39	0.56
4	0.24	0.29	0.49	NA	0.22	0.27
5	0.29	0.43	0.39	0.22	NA	0.56
6	0.48	0.40	0.56	0.27	0.56	NA

NA, not applicable.

Multiusers κ (95% CI) was 0.39 (0.30–0.49), and multiusers κ' was 0.64 (0.55–0.73). Assignment of fibrosis scores by the 15 possible pairings of pathologists to sections stained with picrosirius red is presented in Table 4. The pairs of pathologists were in complete agreement 53.0% of the time, differed by 1 score level 42.3% of the time, and differed by >1 score level 4.7% of the time. There was a significant difference in fibrosis scores assigned to picrosirius red-stained sections by the 6 pathologists ($P < .0001$), and significant differences were found between scores for 2 of 15 possible pathologist pairings.

There was no significant difference median scores assigned by the 6 pathologists for fibrosis between contiguous H&E and picrosirius red-stained sections ($P = .248$).

Necroinflammatory Activity (Grade)

Agreement between the pairs of pathologists assigning scores for necroinflammatory activity is presented in Table 5. The median (minimum–maximum) κ for each pair was 0.19 (–0.03–0.40). Multiusers κ (95% CI) for assessment of necroinflammatory activity was 0.16 (0.10–0.23), and multiusers κ' was 0.43 (0.32–0.55). Assignment of necroinflammatory scores by the 15 possible pairings of pathologists is presented in Table 6. The pairs of pathologists were in complete agreement 34.1% of the time, differed by 1 score level 46.7% of the time, and differed by >1 score level 19.2% of the time. There was a significant difference in necroinflammatory scores assigned to H&E sections by the 6 pathologists ($P < .0001$), and significant differences were found between scores for 7 of 15 possible pathologist pairings.

Discussion

We found fair agreement among veterinary pathologists using a previously published scheme to score hepatic fibrosis in dogs using H&E and picrosirius red-stained sections with κ of 0.35 and 0.39, respectively. These findings are comparable to results of a study

Table 4. Assignment of fibrosis scores by pathologist pairings for picrosirius red-stained sections.

Score	Absent	Mild	Moderate	Marked	Very Marked
Absent	5.2%	NA	NA	NA	NA
Mild	9.6%	9.7%	NA	NA	NA
Moderate	1.7%	14.0%	18.7%	NA	NA
Marked	0.3%	1.6%	11.2%	3.7%	NA
Very Marked	0.0%	0.0%	1.1%	7.5%	15.7%

NA, not applicable.

The percentages represent the frequency at which the 15 possible pathologist pairings assigned fibrosis scores to the sections. The pairs of pathologists were in complete agreement 53.0% of the time (dark gray), differed by 1 score level 42.3% of the time (light gray), and differed by >1 score level 4.7% of the time (white).

Table 5. Pairwise comparison of kappa statistics for the assessment of necroinflammatory activity.

Observer	1	2	3	4	5	6
1	NA	-0.03	0.23	0.14	0.19	0.13
2	-0.03	NA	0.06	0.26	0.08	0.06
3	0.23	0.06	NA	0.11	0.22	0.26
4	0.14	0.26	0.11	NA	0.20	0.37
5	0.19	0.08	0.22	0.20	NA	0.40
6	0.13	0.06	0.26	0.37	0.40	NA

NA, not applicable.

using the Ishak scheme, in which pairwise κ ranged from 0.26 to 0.47, indicating fair to moderate agreement.¹⁵ Although the results from our study and those of the study in humans were similar, they cannot be compared directly because we used a 5-point scheme rather than the original 7-point scheme. When the pathologists did not completely agree, they often assigned scores that deviated only by 1 level, and they deviated only by >1 level 15% and 5% of the time for H&E and picosirius red-stained sections, respectively. This partial agreement was apparent when κ' was calculated. Weighted kappa statistics for H&E and picosirius red-stained sections were 0.59 and 0.64, indicating moderate and good agreement, respectively. In the aforementioned study using the Ishak scheme to score hepatic fibrosis in human patients, pairwise κ' ranged from 0.57 to 0.69, indicating moderate to good agreement.¹⁵ Again, the results are similar to those from our study, but direct comparisons cannot be made.

Interobserver agreement using this scoring scheme to assess hepatic fibrosis in dogs was suboptimal, but a high level of partial agreement was found. There was a significant difference in the fibrosis scores assigned to both the H&E and picosirius red-stained sections by the 6 pathologists. This suggests that some of the disagreement observed was caused by systematic differences in the way the individual pathologists assigned scores. Therefore, it may be possible to improve the level of agreement among pathologists if the descriptions for each score are clarified to encompass differences in interpretation (ie, by defining the subjective descriptors mild, moderate and marked that this scheme uses to define some scores), if the pathologists receive

more training on how to apply the scoring system, or both. Agreement as assessed by κ tends to be higher for histological scoring systems with fewer levels.⁴ For example, in a study of humans evaluating fibrosis using the 4-level METAVIR scoring scheme, κ was reported to be 0.59, indicating moderate agreement.¹⁸ Thus, it may be advantageous to simplify the system we evaluated to a 4-level scale (absent, mild, moderate, and marked).

A multiuser κ of 0.19 indicates poor agreement among pathologists scoring necroinflammatory activity. In human medicine, it also has proven more difficult to develop a grading scheme for hepatic necroinflammatory activity that has acceptable interobserver agreement than 1 for fibrosis.^{4,15} This may be because of the complexity and subjectivity of the histological grading schemes, which must take into consideration a diverse and difficult to standardize set of features, including interface hepatitis, focal necrosis, confluent necrosis, and portal inflammation. The multiuser κ' for necroinflammatory activity was 0.43, indicating moderate interobserver agreement. This reflected the finding that there was often partial agreement between observers. Indeed, the scores assigned by pairs of pathologists only deviated by >1 level 19% of the time. In a study of humans using the Ishak scheme, pairwise κ for the different components of necroinflammatory activity was reported to range from 0.11 to 0.41, indicating poor to moderate agreement, whereas κ' ranged from 0.19 to 0.53, indicating poor to moderate agreement.¹⁵ However, because the scheme we used in this study was a modification of the original Ishak scheme, direct comparisons cannot be made.

Because of this suboptimal interobserver agreement, before use in a clinical setting it would be beneficial to simplify the hepatic necroinflammatory activity scoring scheme that we used. One way to do this would be by collapsing it to 4 levels such as in the METAVIR scheme used to assess chronic hepatitis in humans.¹⁸ It also may be beneficial to score the components of hepatic necrosis and inflammation separately. A significant difference was found in the necroinflammatory scores assigned by the 6 pathologists. As discussed for the scoring of fibrosis, it may be possible to improve the level of agreement among pathologists if the descriptions for each score are clarified, if the pathologists

Table 6. Assignment of necroinflammatory scores by pathologist pairings.

Score	Absent	Slight	Mild	Moderate	Marked	Very Marked
Absent	13.2%	NA	NA	NA	NA	NA
Slight	19.3%	9.1%	NA	NA	NA	NA
Mild	6.1%	11.1%	4.9%	NA	NA	NA
Moderate	2.7%	3.6%	8.7%	3.6%	NA	NA
Marked	0.5%	0.7%	2.3%	4.7%	1.5%	NA
Very marked	0.3%	0.5%	0.7%	1.9%	2.9%	1.9%

NA, not applicable.

The percentages represent the frequency at which the 15 possible pathologist pairings assigned scores to the sections. The pairs of pathologists were in complete agreement 34.1% of the time (dark gray), differed by 1 score level 46.7% of the time (light gray), and differed by >1 score level 19.2% of the time.

receive more training on how to apply the scoring system, or both.

We hypothesized that fibrosis would be easier to detect on picosirius red-stained sections and that these sections would be assigned a higher fibrosis score, but there was no significant difference in median fibrosis scores assigned to H&E and picosirius red-stained sections. The agreement associated with the evaluation of fibrosis from sections stained with picosirius red was slightly higher than that from H&E-stained sections, κ statistics of 0.35 (0.26–0.44) and 0.39 (0.22–0.56), respectively. This difference should be interpreted cautiously because there was considerable overlap between CI. Taken together, these findings did not indicate a clear objective benefit of staining sections with picosirius red. However, some of the coauthors still prefer this stain for the assessment of hepatic fibrosis.

Our study had several limitations. The kappa statistic is commonly used to assess interobserver agreement in biomedical research, but some authors have criticized its use. One concern is that κ and κ' are dependent upon the distribution of severity among the cases.¹⁹ The cases in this study were selected primarily to represent a wide range of fibrosis scores, and although a wide range of severity of necroinflammatory activity scores was represented, the distribution among median scores was not even. Therefore, the absolute values of κ and κ' for the assessment of interobserver agreement associated with histological scoring of this variable should be interpreted cautiously. Interpretation guidelines and weighting schemes used to calculate the κ' also have been criticized as being too subjective.¹⁹ Because of the limitations of κ , we also expressed results as the frequency of different levels of agreement between pathologists. Another potential limitation is the use of material from healthy dogs and dogs with congenital portosystemic shunts. This was done to ensure that patients with no or mild fibrosis were included, and we believe doing so is unlikely to have influenced our main conclusion that there was disagreement among pathologists. Only 1 of the 6 pathologists had previously used a grading scheme for assessment of hepatic biopsy samples and introducing pictorial templates and example cases may help to decrease the interobserver variability documented in our study. As previously discussed, a limitation of the scoring scheme that we used was the use of subjective descriptors such as mild to define some of its score levels. Although 3 pathologists evaluated biopsy samples based upon whole-slide digital images and 3 used conventional light microscopy, this was not thought to affect the results because previous studies indicated excellent agreement between the 2 modalities.^{20–23} The effect of pathologist experience and specialization on interobserver agreement has been found to be important in previous studies,¹⁶ but our study did not evaluate this factor or consider differences in biopsy technique, both of which are worthy of further investigation. It also would be useful to determine the intra-observer agreement for pathologists using this scheme.

In conclusion, use of this scoring scheme resulted in fair interobserver agreement when pathologists assessed hepatic fibrosis in dogs and poor agreement when they assessed hepatic necroinflammatory activity. This sub-optimal agreement is concerning and to decrease this variability ideally >1 pathologist should evaluate each section. A simplified scoring scheme with fewer, more clearly defined levels may improve interobserver agreement. Additionally, fibrosis scores and interobserver agreement were similar for serial sections of liver stained with H&E and picosirius red.

Footnotes

- ^a Picosirius red stain kit, PolySciences, Warrington, PA
^b Weigert's hematoxylin stain kit, Polysciences, Warrington, PA
^c Nanozoomer 2.0-HT, Hamamatsu Photonics, Hamamatsu City, Shizuoka Pref., Japan
^d Stata v12, StataCorp, College Station, TX
^e Prism v5, GraphPad, La Jolla, CA
-

Acknowledgments

Grant support: None

Conflict of Interest Declaration: Authors are affiliated with entities that provide histological assessment of canine liver tissue on a fee for service basis. This did not lead to any conflict of interest or influenced the collection or interpretation of results. The pathologists were blinded to the identity of the slides, each others scores, and image analysis results.

Off-label Antimicrobial Declaration: Authors declare no off-label use of antimicrobials.

References

1. Knodell RG, Ishak KG, Black WC, et al. Formulation and application of a numerical scoring system for assessing histological activity in asymptomatic chronic active hepatitis. *Hepatology* 1981;1:431–435.
2. Ishak K, Baptista A, Bianchi L, et al. Histological grading and staging of chronic hepatitis. *J Hepatol* 1995;22:696–699.
3. Bedossa P, Poinard T. An algorithm for the grading of activity in chronic hepatitis C. The METAVIR Cooperative Study Group. *Hepatology* 1996;24:289–293.
4. Goodman ZD. Grading and staging systems for inflammation and fibrosis in chronic liver diseases. *J Hepatol* 2007;47:598–607.
5. Poldervaart JH, Favier RP, Penning LC, et al. Primary hepatitis in dogs: A retrospective review (2002–2006). *J Vet Intern Med* 2009;23:72–80.
6. Fieten H, Biourge VC, Watson AL, et al. Nutritional management of inherited copper-associated hepatitis in the Labrador retriever. *Vet J* 2014;199:429–433.
7. Fieten H, Dirksen K, van den Ingh TS, et al. D-penicillamine treatment of copper-associated hepatitis in Labrador retrievers. *Vet J* 2013;196:522–527.
8. van den Ingh TS, Van Winkle T, Cullen JM, et al. Morphological classification of parenchymal disorders of the canine and feline liver: Hepatocellular death, hepatitis, and cirrhosis-2 (updated

version). In: WSAVA Standards for Clinical and Histological Diagnosis of Canine and Feline Liver Diseases. Society of Comparative Hepatology. Available at: <http://www.vetvisuals.com/lms/moodle/mod/book/view.php?id=1001&chapterid=52859>. Accessed December 12, 2016.

9. Willard MD, Jergens AE, Duncan RB, et al. Interobserver variation among histopathologic evaluations of intestinal tissues from dogs and cats. *J Am Vet Med Assoc* 2002;220:1177–1182.

10. Willard MD, Moore GE, Denton BD, et al. Effect of tissue processing on assessment of endoscopic intestinal biopsies in dogs and cats. *J Vet Intern Med* 2010;24:84–89.

11. Cole TL, Center SA, Flood SN, et al. Diagnostic comparison of needle and wedge biopsy specimens of the liver in dogs and cats. *J Am Vet Med Assoc* 2002;220:1483–1490.

12. Kemp SD, Zimmerman KL, Panciera DL, et al. A comparison of liver sampling techniques in dogs. *J Vet Intern Med* 2015;29:51–57.

13. Kemp SD, Zimmerman KL, Panciera DL, et al. Histopathologic variation between liver lobes in dogs. *J Vet Intern Med* 2015;29:58–62.

14. Cohen J. Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull* 1968;70:213–220.

15. Westin J, Lagging LM, Wejstal R, et al. Interobserver study of liver histopathology using the Ishak score in patients with chronic hepatitis C virus infection. *Liver* 1999;19:183–187.

16. Cross SS, Betmouni S, Burton JL, et al. What levels of agreement can be expected between histopathologists assigning cases to discrete nominal categories? A study of the diagnosis of hyperplastic and adenomatous colorectal polyps. *Mod Pathol* 2000;13:941–944.

17. Gadbois J, d’Anjou MA, Dunn M, et al. Radiographic abnormalities in cats with feline bronchial disease and intra- and interobserver variability in radiographic interpretation: 40 cases (1999–2006). *J Am Vet Med Assoc* 2009;234:367–375.

18. Rousselet MC, Michalak S, Dupre F, et al. Sources of variability in histological scoring of chronic viral hepatitis. *Hepatology* 2005;41:257–264.

19. Crewson PE. Fundamentals of clinical research for radiologists—reader agreement studies. *Am J Roentgenol* 2005;184:1391–1397.

20. Houghton JP, Ervine AJ, Kenny SL, et al. Concordance between digital pathology and light microscopy in general surgical pathology: A pilot study of 100 cases. *J Clin Pathol* 2014;67:1052–1055.

21. Campbell WS, Lele SM, West WW, et al. Concordance between whole-slide imaging and light microscopy for routine surgical pathology. *Hum Pathol* 2012;43:1739–1744.

22. Al-Janabi S, Huisman A, Vink A, et al. Whole slide images for primary diagnostics of gastrointestinal tract pathology: A feasibility study. *Hum Pathol* 2012;43:702–707.

23. Al-Janabi S, Huisman A, Vink A, et al. Whole slide images for primary diagnostics in dermatopathology: A feasibility study. *J Clin Pathol* 2012;65:152–158.

Appendix: Histological grading and staging system for canine chronic hepatitis

Necroinflammatory Activity (Grade)			
Grade	Periportal or Periseptal Interface Hepatitis	Focal Lytic Necrosis	Confluent Necrosis
Absent (0)	Absent	Absent	Absent
Slight (1)	Very mild	1 focus per 10× obj.	Absent
Mild (2)	Mild	2–4 foci per 10× obj.	Absent
Moderate (3)	Moderate	5–10 foci per 10× obj.	Absent
Marked (4)	Marked	>10 foci per 10× obj. and/or →	Confluent or bridging necrosis
Very marked (5)	Marked	>10 foci per 10× obj. and/or →	Bridging or panacinar/multiacinar necrosis

Degree of Fibrosis (Stage)			
Stage	Fibrosis	Bridging Fibrosis	Bridging Fibrosis with Nodule Formation
Absent (0)	Absent	Absent	Absent
Mild (1)	Mild fibrous expansion (periportal or central)	Absent	Absent
Moderate (2)	Moderate fibrous expansion	Some bridging fibrosis (PP, CC, or PC)	Absent
Marked (3)	Marked fibrous expansion	Marked bridging fibrosis (PP, CC, or PC)	Absent
Very marked (4)	Marked fibrous expansion	Marked bridging fibrosis (PP, CC, or PC)	Present

PP, portal–portal; CC, central–central; PC, portal–central.

A focus was defined as a discrete area of inflammation or necrosis and interface hepatitis is defined as inflammation and erosion of the hepatic parenchyma at its junction with portal tracts or fibrous septa.

Reproduced with permission from: van den Ingh et al.⁸