Intra- and interobserver variability of board-certified veterinary radiologists and veterinary general practitioners for pulmonary nodule detection in standard and inverted display mode images of digital thoracic radiographs of dogs

David J. Reese, DVM, DACVR; Eric M. Green, DVM, DACVR; Lisa J. Zekas, DVM, DACVR; Jane E. Flores, DVM; Lawrencen N. Hill, DVM, DABVP; Matthew D. Winter, DVM, DACVR; Clifford R. Berry, DVM, DACVR; Norman Ackerman,† DVM, DACVR

Objective—To determine intra- and interobserver variability of 2 veterinary radiologists and 2 veterinary general practitioners for detection of pulmonary nodules in standard and inverted (reversed grayscale) displays of digital thoracic radiographs of dogs.

Design—Evaluation study.

Sample—114 sets of 3-view (right lateral, left lateral, and ventrodorsal or dorsoventral views) digital thoracic radiographs from 114 dogs.

Procedures—2 experienced board-certified veterinary radiologists and 2 experienced veterinary general practitioners individually evaluated 114 randomized sets of radiographs. Pulmonary nodules were present in radiographs of 60 of 114 dogs. Each reviewer examined all images in standard or inverted display mode and scored nodule detection on a confidence scale of 1 to 5. After ≥ 2 months, the same individuals evaluated the same images in the remaining display mode. Intraobserver agreement for each display mode was determined via a $\kappa$ statistic; results between the 2 groups of reviewers were compared via receiver operator curve analysis.

Results—There was no significant intraobserver variability in pulmonary nodule detection between the 2 display modes. Detection accuracy for board-certified radiologists was significantly greater than that of veterinary general practitioners for both display modes. Nearperfect intraobserver agreement was detected between the 2 display modes for board-certified radiologists, whereas moderate to slight intraobserver agreement was detected for the veterinary general practitioners.

Conclusions and Clinical Relevance—Detection of pulmonary nodules in digital thoracic radiographs was comparable, whether a standard or inverted mode was used for evaluations. However, the board-certified radiologists had greater detection accuracy than did veterinary general practitioners. (J Am Vet Med Assoc 2011;238:998–1003).

Pulmonary nodule detection for metastatic disease screening poses a challenge in human and veterinary medicine. Computed tomography has been proven superior to conventional radiography for the detection of pulmonary nodules in humans.1–4 Currently, a 3-view (left lateral, right lateral, and ventrodorsal or dorsoventral) radiographic evaluation of the thorax is considered the standard of care for evaluation of pulmonary metastasis and other lung lesions in veterinary patients.5,6 Studies5–11 in dogs and cats have estimated the sensitivity of radiographic detection of metastatic nodules to range from 64% to 97%. The staging of malignant disease is partially based on the extent of metastasis to distant tissues, and detection of pulmonary nodules can greatly influence treatment decisions for the clinician and pet owner.

The implementation of DR in veterinary medicine has provided the opportunity to greatly improve patient imaging efficiency and distribution of images for interpretation.12–15 Postprocessing algorithms for viewing digital radiographs allow the reviewer to manipulate an image to individual preferences.16 Any postprocess-
The use of DR is well-established in humans, and human clinical trials have focused on the use of post-processing manipulation for the detection of pulmonary nodules. One form of post-processing that has been evaluated in humans is reversal of the image grayscale, which produces a positive or inverted image in which the background contrast is shifted so that areas that originally were dark (eg, aerated lung) become bright on the inverted image. Thus, bone appears black, and air spaces appear white. It has been postulated that nodules that are superimposed against a darker background, this type of inverted display may aid in the detection of pulmonary nodules. This display mode was originally thought to provide an easier means of detecting structures that have low contrast in humans; however, in a study performed in 1983, this was described as a subjective finding and was not critically evaluated.

The benefit of specific methods of radiographic post-processing has not been well documented in veterinary medicine; most information on this subject is anecdotal. Herein, we report results from studies in dogs.

The purpose of the study reported here was to determine intra- and interobserver variability of 2 board-certified veterinary radiologists and 2 veterinary general practitioners for pulmonary nodule detection in SDIs and IDIs of digital thoracic radiographs of dogs as a means of evaluating the utility of these methods. The null hypotheses were that no intraobserver variability would be determined for detection of pulmonary nodules between the 2 display modes, and that no interobserver variability would be determined for detection of pulmonary nodules in SDI and IDI modes.

Materials and Methods

Image selection—Digital images created by means of CR (n = 62) or DR (52) from October 1, 2005, through March 31, 2007, were obtained from medical records of the University of Florida Veterinary Medical Center and the College of Veterinary Medicine at The Ohio State University, respectively. Each set of images consisted of a 3-view (right lateral, left lateral, and ventrodorsal or dorsoventral) series of thoracic radiographs. Keyword searches of specific terms (nodule, metastasis, and structured interstitial) were performed to identify records of dogs with pulmonary nodules. Inclusion of these images in the study required the presence of 1 to 6 soft tissue pulmonary nodules < 3 cm in size. A total of 60 sets of 3-view thoracic radiographs (from 60 dogs) were identified from both databases for inclusion in the study. Additionally, 54 sets of 3-view thoracic radiographs (from 54 size- and age-matched dogs) that were interpreted to be free of pulmonary nodules were included in the study. Each set of images was evaluated by one of the authors (DJR) for acceptable diagnostic quality, and the order in which each set of 3 radiographs (with or without pulmonary nodules) was presented to individual reviewers was randomly assigned according to the dogs’ birthdates placed in chronological order. This order was then followed or changed to reverse chronological order for each reviewer by one of the authors (DJR). The author that performed the randomization was therefore not aware of the specific sequence supplied to each reviewer.

Evaluation of DR images—All 114 sets of images were independently evaluated by 4 reviewers. Two were board-certified veterinary radiologists (LJZ and EMG) with 8 years and 5 years of experience as American College of Veterinary Radiologists diplomates, respectively, and 2 were veterinary general practitioners in small animal practice (JEF and LNH) with 2 years and 11 years of academic practice experience and 9 years and 3 years of private practice experience, respectively. All images were examined in a designated radiology interpretation room and viewed by use of standard computer workstations with Digital Imaging and Communications in Medicine (ie, DICOM) viewing software and 2 3-megapixel grayscale monitors. To reduce reviewer fatigue, the reviewers were instructed to examine images in 30-minute sessions with 10- to 15-minute breaks between sessions. The reviewers were unaware of patient histories and the total numbers of dogs that did or did not have pulmonary nodules.

Each set of 3-view thoracic radiographs was evaluated with 2 separate display settings: standard (in which bones appeared white) and inverted (in which bones appeared black). Interpretation sessions for the images viewed in SDI and IDI modes were ≥ 2 months apart. The image window-level (ie, brightness and contrast) display was selected as linear for images viewed in either mode. Reviewers were allowed to change brightness and contrast and to pan and enlarge the image by use of the DICOM viewing software. However, to minimize potential influences on the image display, they were not permitted to use other methods of image enlargement (eg, the magnifying glass tool) or to manipulate the image in any other way. The reviewers scored their assessments for the presence of pulmonary nodules on a 5-category confidence rating scale (1 = definitely absent; 2 = probably absent; 3 = indeterminate; 4 = probably present; and 5 = definitively present). For each display mode, scores were marked on an individual examination sheet for each of the 114 image sets. Additionally, the locations of any nodules detected were marked on a diagram of the thorax so that the affected lung lobe was identified by the reviewer.

After all reviewers had completed the 2 viewing sessions, the images were assessed by 3 investigators (board-certified veterinary radiologists [DJR, MDW, and CRB]). Each image was evaluated in the SDI display mode by use of a 4-panel, 3-megapixel grayscale workstation; the presence of 1 or more pulmonary nodules or absence of nodules was determined by consensus. This was determined to be the gold standard against which the reviewers’ scores would be compared (in both SDI and IDI displays). The number of nodules present was determined if applicable, and the size of each nodule on the radiograph was measured with digital calipers included in the DICOM viewing software.

Statistical analysis—Sensitivity and specificity of nodule detection were determined for each reviewer for
each data set on the basis of comparison with consensus panel evaluations, and ROC curves for pulmonary nodule detection were plotted and analyzed. Area under the ROC curve was calculated for the 2 radiologists and the 2 veterinary general practitioners for images in SDI and IDI modes. Statistical analysis of the ROC curves was used to evaluate detection differences between display modes for individual reviewers (intraobserver variation) as well as differences between radiologists and general practitioners (interobserver variation). This was determined by use of the mean areas under the ROC curves for the radiologists and the general veterinary practitioners, compared via a modified Wilcoxon rank sum statistic. A 2 statistic was determined to evaluate the level of intraobserver agreement between the 2 display modes for detection of nodules (confidence scale assessments > 4 were categorized as reporting nodules present, and those ≤ 3 were categorized as reporting nodules absent). The level of agreement was assessed as follows: \( \kappa < 0 \), poor agreement; \( \kappa = 0 \) to 0.20, slight agreement; \( \kappa = 0.21 \) to 0.40, fair agreement; \( \kappa = 0.41 \) to 0.60, moderate agreement; \( \kappa = 0.61 \) to 0.80, substantial agreement; and \( \kappa = 0.81 \) to 1.0, near-perfect agreement. A value of \( P < 0.05 \) was considered significant for all statistical tests. All statistical analyses were performed by use of commercially available software.

Results

A total of 114 sets of 3-view thoracic digital radiographs were selected for evaluation. Pulmonary nodules were determined to be present in 60 sets of radiographs (25 sets of DR images and 35 sets of CR images) and absent in 54 sets (27 sets each of DR and CR images). The consensus of the 3 investigating radiologists (DJR, MDW, and CRB) for image sets with detectable pulmonary nodules was as follows: 26 image sets had 1 nodule, 11 image sets had 2 nodules, 5 image sets had 3 nodules, 4 image sets had 4 nodules, and 14 image sets had ≥ 5 nodules (each; Figures 1 and 2). The pulmonary nodule sizes ranged from 3 to 25 mm, as measured in the consensus SDI viewing.

Figure 1—Left lateral radiographic view of the thorax of a dog, shown in SDI (A) and IDI (B) modes. Well-defined, opaque soft tissue nodules are present in the right cranial lung lobe, superimposed over the middle aspect of the fourth rib (arrows), and within the ventral periphery of the right middle lung lobe (arrowheads).

Figure 2—Right lateral radiographic view of the thorax of a dog, shown in SDI (A) and IDI (B) modes. An ill-defined, opaque soft tissue nodule is present in the seventh intercostal space, dorsal to the caudal vena cava (arrows). This nodule was detected in both display modes by the 2 radiologists that participated in the study but was not reported by the 2 veterinary general practitioners.
The areas ± SE under the ROC curves were summarized for each of the 4 reviewers (Table 1). There was no significant difference in the accuracy of pulmonary nodule detection by radiologists, whether the radiographs were viewed in SDI or IDI modes. There was also no significant difference in the accuracy of pulmonary nodule detection between the 2 display modes for either of the general practitioners.

The mean areas ± SE under the ROC curve for detection of pulmonary nodules in the SDI mode were significantly \( (P < 0.001) \) different between radiologists \( (0.91 ± 0.029) \) and general practitioners \( (0.69 ± 0.03) \); \( \text{Figure 3} \). In the IDI mode, there was also a significant \( (P < 0.001) \) difference in mean areas ± SE under the ROC curve between radiologists \( (0.90 ± 0.02) \) and general practitioners \( (0.62 ± 0.04) \).

Each radiologist had near-perfect intraobserver agreement between the IDI and SDI modes for detection of pulmonary nodules \( (\kappa = 0.82 \text{ and } 0.86, \text{ respectively}) \). Intraobserver agreement between the 2 display modes was moderate \( (\kappa = 0.48) \) for 1 of the 2 general practitioners and slight \( (\kappa = 0.20) \) for the other. Post hoc power analysis revealed a low power associated with each individual reviewer \( \text{radiologist 1, 0.14; radiologist 2, 0.05; general practitioner 1, 0.34; and general practitioner 2, 0.13).} \)

Discussion

In the study reported here, pulmonary nodule detection by veterinary radiologists and veterinary general practitioners in 3-view digital thoracic radiographs of dogs was not improved by examination of IDIs, compared with results for SDIs. This was similar to findings reported in human studies \( ^{17,21} \) in which inverting the image display had no effect on detection of pulmonary nodules. In another study \( ^{20} \), detection of subtle lung nodules in humans by use of digitized analog radiographs was impaired by inverting the displayed images. Another report \( ^{22} \) in which investigators assessed a variety of subtle thoracic abnormalities in humans (eg, pulmonary nodules, pneumothorax, bone lesions, unstructured interstitial pulmonary pattern) indicated that images viewed in IDI mode were significantly inferior to those viewed in SDI mode.

In the present study, the area under the ROC curve for pulmonary nodule detection appeared to be smaller for radiographs viewed in IDI mode than for those
viewed in SDI mode for all examiners; however, the difference was not significant for any individual reviewer. This suggests that there is no benefit from exclusively viewing thoracic radiographs in the SDI or IDI mode. Alternatively, this finding may reflect the lack of reviewer experience in assessing images in the IDI mode. Inversion of mineral opacities (eg, pulmonary osteomas) may make these more difficult to discern from soft tissue nodules for reviewers unfamiliar with examining images in this mode. Results may be different for examiners who routinely view thoracic radiographs in the IDI mode. Additionally, the number of image sets in the present study was small, which could result in a type II sampling error, so that true intraobserver differences between SDI and IDI modes of viewing could have been missed. In future studies, a sample size of up to 10,000 sets would be needed to detect these differences.

Post hoc power analysis resulted in a lower number for radiologists than for general practitioners because of similar responses between the 2 display modes. An additional source of error is that the radiologists were more familiar with reviewing radiographs at a computer workstation than were the general practitioners. A lower comfort level for reviewing digital radiographs may have contributed to a less-accurate thoracic radiographic review for the detection of pulmonary nodules by general practitioners. To minimize the number of independent variables, reviewers in the present study were also limited to using only the DICOM viewer pan, enlarge, and brightness and contrast tools for image evaluation. This may have had a more negative impact on results for general practitioners than for radiologists and could provide some insight for planning future investigations. Additionally, the number of reviewers was small, with only 2 individuals in each group.

The second null hypothesis was rejected because the overall accuracy of pulmonary nodule detection for general practitioners was significantly \( P < 0.01 \) less than that for radiologists. This difference in performance may indicate that the radiologists were more proficient in pattern recognition, allowing for greater discrimination between pulmonary nodules (ie, true-positive results) and end-on vessels or pulmonary osteomas (false-positive results). The radiologists also may have had a higher threshold for determining the presence of pulmonary nodules, resulting in a smaller number of false-positive results.

Several weaknesses were identified for the study reported here. The presence or absence of nodules was determined by consensus among 3 veterinary radiologists that did not participate in the image review sessions. The presence of nodules was not confirmed by means of necropsy or additional imaging methods (ie, computed tomography). Although histology or necropsy is considered a gold standard, investigators of earlier studies in humans and nonhuman animals used methods similar to those reported here to evaluate thoracic radiographs for the presence of pulmonary lesions. Pulmonary metastasis from a distant neoplastic lesion was not determined in the present study; however, the purpose of the study was to assess reviewers’ abilities to detect pulmonary nodules in digital thoracic radiographs in the 2 display modes and not specifically to determine etiology of pulmonary nodules. The potential benefit of viewing a set of 3-view thoracic radiographs in both display modes during the same review session was also not evaluated.

Images used in the present study were obtained by means of 2 different radiography methods (CR and DR) that involve the use of different pre- and postprocessing algorithms. Statistical analysis was not performed to determine whether reviewers' responses differed between the 2 different imaging methods. Additionally, the 3 investigating radiologists that performed consensus review of the images did not use the same computer workstations used by the 4 reviewers in the study. However, both workstations included DICOM viewers and medical-grade (3-megapixel) monitors. The reviewers (radiologists and practitioners) were asked only to determine the presence or absence of pulmonary nodules and did not report other abnormalities that were present. No time limits were specified for examination of the images, which may have improved nodule detection in both display modes.

On the basis of results of the present study, detection of pulmonary nodules in 3-view digital thoracic radiographs of dogs appears to be comparable between SDI and IDI modes. However, the accuracy of pulmonary nodule detection is affected by the advanced training of the reviewer.


b. Sound-Eklin EDR6, Sound-Eklin Medical Systems Inc, Carlsbad, Calif.
c. eFilm Workstation 2.1, version 2.1.2, Merge Healthcare, Milwaukee, Wis.
d. WIDE IF2103M LCD 3M pixel, Wide Co Ltd, Cheongwon-gun, Chunghw, Korea.
e. Kodak DX Workstation System 5, Carestream Health, Health Imaging Group, Rochester, NY.
g. PASS 2008, NCSS, Kaysville, Utah.

References
From this month’s AJVR

Evaluation of a nine-point body condition scoring system in physically inactive pet cats

Charlotte R. Bjornvad et al

**Objective**—To compare results of body condition scoring by use of a 9-point scale with body composition determined by dual-energy x-ray absorptiometry (DEXA) in indoor-confined neutered domestic shorthair (DSH) pet cats.

**Animals**—72 indoor-confined, adult neutered DSH pet cats (38 females and 34 males).

**Procedures**—All cats underwent a physical examination including assessment of body weight (BW), body condition score (BCS, 1 = emaciated, 5 = ideal, and 9 = grossly obese), and girth. Urinalysis, CBC, and serum biochemical analysis were also performed. After the cats were confirmed healthy, they were anesthetized for body composition measurement via DEXA. Lean body mass, fat mass, and percentage body fat (%BF) were then evaluated.

**Results**—The correlation between %BF and BCS ($r = 0.87$) was superior to the correlations between %BF and BW ($r = 0.74$) and between %BF and girth ($r = 0.78$). Values for %BF differed significantly between all pairs of BCSs except BCSs 8 and 9. Within a BCS, the %BF was similar for male and female cats. The mean %BF for cats with a BCS of 5 was 32, which exceeded the upper reference limit of %BF generally considered ideal (30).

**Conclusions and Clinical Relevance**—The 9-point BCS scale appears useful for assessing %BF in DSH pet cats. Nevertheless, study findings could indicate a need for redefining the ideal BCS for inactive neutered cats to include a BCS of 4. (Am J Vet Res 2011;72:433–437)