Evaluation of intra- and interobserver reliability and image reproducibility to assess usefulness of high-resolution ultrasonography for measurement of anterior segment structures of canine eyes

Ellison Bentley, DVM; Paul E. Miller, DVM; Kathryn A. Diehl, DVM, MS

Objective—To evaluate the usefulness of high-resolution ultrasonography (HRUS) for measurements of anterior segment structures in canine eyes.

Animals—4 clinically normal Beagles.

Procedure—Images were obtained from 8 eyes with a handheld 20-MHz transducer. Eleven anterior segment structures on each image were measured 5 times by 2 independent observers. Coefficients of variation (CVs) for measurements were used to assess intraobserver reliability. Interobserver reliability was assessed by comparing measurements obtained by the 2 observers from the same images. Five images were sequentially obtained from 2 locations (ie, superior and temporal) to evaluate image reproducibility. Anterior segment structures were measured once on each image; image reproducibility was assessed by use of the CV for each parameter measured. Imaging location was assessed by comparison of CV for measurements from each location.

Results—CVs were < 10% for observer A for all measurements except the ciliary cleft area (11.63%). The CVs were > 10% for observer B for measurements of the angle recess area (18.51%) and ciliary cleft width (17.44%) and area (16.01%). Significant differences in measurements between observers were found for 5 of 11 anterior segment structures. Imaging the superior aspect of the globe provided the most reproducible images, although image reproducibility was still somewhat variable, with the highest and lowest CVs for measurements of 33.01% and 11.32%, respectively, in the superior position.

Conclusions and Clinical Relevance—High-resolution ultrasound images can be used to reliably measure various anterior segment structures. Clinically relevant findings in the anterior segment of canine eyes may be detectable by use of HRUS. (*Am J Vet Res* 2005;66:1775–1779)

High-resolution ultrasonography (HRUS) has been recently described as a diagnostic tool in veterinary ophthalmology.[1] High-resolution ultrasonography allows noninvasive observation of living tissues in vivo at resolutions of 20 to 80 µm, similar to a low-power histologic view. Unlike histologic evaluation, however, HRUS allows the tissue to be viewed without the distortion associated with fixation and tissue processing and also allows sequential examination of the eye. Tissue penetration of 5 to 10 mm makes this an ideal modality for examination of the anterior segment of the eye.[2,3] High-resolution ultrasonography is useful for evaluation of tumors and cysts of the anterior segment, evaluation of the cornea and sclera, assessment of the lens, and differentiation of various forms of glaucoma.

Anterior segment structures of the eye that are typically obscured from view when normal examination techniques, such as slit-lamp biomicroscopy and gonioscopy are used, or by opacities in the cornea, aqueous humor, or vitreous humor can be examined by use of HRUS. High-resolution ultrasonography can be used to critically evaluate normal anatomic relationships between anterior segment structures and those that occur in diseased states, thereby giving insight into the underlying pathophysiologic characteristics of various anterior segment disorders. High-resolution ultrasonography can also be used to quantitatively monitor disease processes, such as tumor progression or response to treatment. Quantitative evaluation of high-resolution ultrasound images would also allow for evaluation of drug effects on anterior segment structures to determine mechanisms of action that are often difficult to otherwise elucidate.

For HRUS to be fully used as a clinical and research tool, users must be able to obtain reliable and reproducible quantitative measurements of structures and anatomic relationships. One step toward achieving this goal is to obtain reproducible images. Image reproducibility is influenced by the transducer angle, plane of section, patient cooperation, variation in the technique of individual users, and physiologic changes such as pupil size and accommodation. Another potential barrier to the goal of quantitative measurement is variation in values obtained by observers. Although many studies[4] report documenting quantitative changes measured by ultrasound biomicroscopy in response to disease processes or administered drugs, few studies[5-7] in physician-based ophthalmology or veterinary ophthalmology have examined observer reliability and image reproducibility. These previous studies[5-8] have established that intraobserver reliability is higher than interobserver reliability and that reproducibility and reliability are best achieved with a single observer measuring

Received November 10, 2004.
Accepted February 1, 2005.
From the Department of Surgical Sciences, School of Veterinary Medicine, University of Wisconsin, Madison, WI 53706-1102.
Supported by the Steven Bechtel Foundation.
Presented in part at the 33rd Annual Meeting of the American College of Veterinary Ophthalmologists and the Association of Research and Vision in Ophthalmology, Ft Lauderdale, FL 2003.
Address correspondence to Dr. Bentley.

well-defined ocular structures. The purpose of the study reported here was to evaluate intra- and interobserver reliability and image reproducibility of high-resolution ultrasound images in dogs to establish quantitative, objective parameters to monitor disease processes and to facilitate research to elucidate the pathophysiologic characteristics of anterior segment disorders and the mechanism of action of ophthalmic treatments. For the purpose of this study, reliability is defined as the ability of observers to obtain similar measurements from the same images, whereas image reproducibility is defined by the ability of an observer to obtain similar images from the same eye in a consistent fashion, as assessed by measurements of anterior segment structures.

Materials and Methods

Procedure—Procedures were approved by the Animal Care and Use Committee of the University of Wisconsin-Madison and conformed to the Association for Research in Vision and Ophthalmology's statement for the use of animals in ophthalmic and vision research. All images were obtained by use of a 20-MHz probe attached to a commercially available B-scan ultrasound machine.[a] Dogs were examined with manual restraint in sternal recumbency after topical administration of 0.5% proparacaine hydrochloride.[b] All examinations were performed in the same room under the same lighting conditions. All images were obtained with the probe in the longitudinal position, meaning that the scan plane was perpendicular to the limbus. Eleven parameters adapted for the canine eye from previous studies[4,6,7] in humans were evaluated (**Figures 1 and 2**) as follows: 1) **angle recess area (ARA)**, measured from the end of Descemet's membrane to the iris on a line perpendicular to Descemet's membrane, then following the surface of the iris to the trabecular meshwork back up to the end of Descemet's membrane; 2) **scleral thickness (ST)**, peripheral ST measured from the end of Descemet's membrane to the outer surface of the sclera on a perpendicular line; 3) **corneal thickness (CT)**, peripheral CT measured from the epithelial surface to the posterior aspect of Descemet's membrane; 4) iris thickness, peripheral iris thickness measured from the base of the trabecular meshwork to the posterior aspect of the pigmented epithelium; 5) maximum length of the ciliary cleft; 6) **maximum width of the ciliary cleft (CCW)**; 7) **area of the ciliary cleft (CCA)**; 8) **trabecular meshwork-iris distance (TID)**, measured as the shortest distance from the end of Descemet's membrane directly to the iris surface; 9) iridociliary process distance, measured from the end of Descemet's membrane directly to the superior surface of a ciliary process; 10) angle opening distance at 500 mm measured 500 mm anteriorly from the end of Descemet's membrane, then on a perpendicular line to the iris; and 11) iris-lens contact, measured from the iris pigmented epithelium at the pupillary margin to the point where the iris no longer contacts lens capsule.

Analysis of intra- and interobserver reliability—Eight eyes of 4 normal young adult laboratory Beagles were used in this study. Findings on ophthalmic examinations, including slit-lamp biomicroscopy and indirect ophthalmoscopy, were normal. All eyes underwent HRUS as already described. Images were imported into a computerized imaging program.[c] For intraobserver reliability, 2 observers measured all anterior segment structures on 1 image from each eye on 5 occasions, a minimum of several hours apart. Observers were unaware of results obtained by the other observer. **Coefficients of variation (CVs)** were calculated for each measurement for each observer. A CV of < 10% for a measurement was considered to indicate good reliability.[7] Observer A had more experience reading high-resolution ultrasound images than observer B, who was trained to measure images at the time of this study.

Interobserver reliability was obtained by comparing measurements obtained by the 2 observers measuring the same images. Reliability was assessed use of a 2-way ANOVA.[6,7] Values of $P < 0.05$ were considered significant, demonstrating a significant difference between measurements obtained by the 2 observers and, hence, poor interobserver reliability.

Analysis of image reproducibility—Ultrasonography was performed as already described on the same 4 clinically normal dogs. Five images sequentially obtained from the 12 o'clock location (denoted as superior) were compared with images taken at either the 9 o'clock (right eye) or 3 o'clock (left eye, both denoted as temporal) locations in both eyes of 4 dogs. These positions were chosen because these regions of the eye are easiest to access in conscious dogs. The lighting conditions and position of the dog were held constant throughout the examination. The same 11 parameters were measured on the images as already described, and CVs were calculated for each parameter at each imaging position. The same operator obtained the images and the same observer measured the images for this portion of the study. The CVs for measurements were then compared between the 2 positions to assess image reproducibility at both locations.

Results

Intra- and interobserver reliability—For observer A, CVs were < 10% for all parameters measured except the CCA (CV, 11.63%). For observer B, CVs were < 10% for all parameters measured except the ARA (CV, 18.51%) and CCW (CV, 17.44%). Interobserver reliability was lower, with significant differences in values of the 2 observers for 5 (ie, ARA, ST, CT, CCW, and CCA) of the 11 anterior segment structures measured (**Table 1**).

Image reproducibility—Images obtained from the superior position had a lower CV for 7 of the 11 parameters measured, compared with those of images obtained from the temporal positions, although the dif-

Table 1—Intra- and interobserver reliability of measurements of high-resolution ultrasound images of the canine eye.

Anterior segment structures	Intraobserver CV (%)		
	Observer A	Observer B	Interobserver *P* value*
ARA	8.458	18.51†	< 0.05
ST	4.214	10.00	< 0.05
CT	4.343	6.207	< 0.05
IT	8.66	6.845	0.203
CCL	6.664	8.097	0.122
CCW	9.623	17.44†	< 0.05
CCA	11.63†	16.01†	< 0.05
TID	6.256	9.491	0.128
IPD	4.721	4.634	0.244
AOD	4.985	6.924	0.063
ILC	4.661	7.072	0.3808

*$P < 0.05$ suggests poor interobserver reliability. †Coefficient of variation was > 10%.

CV = Coefficient of variation. ARA = Angle recess area. ST = Scleral thickness. CT = Corneal thickness. IT = Peripheral iris thickness. CCL = Maximum length of ciliary cleft. CCW = Maximum width of ciliary cleft. CCA = Area of ciliary cleft. TID = Trabecular meshwork-iris distance. IPD = Iridociliary process distance. AOD = Angle opening distance at 500 µm. ILC = Iris-lens contact.

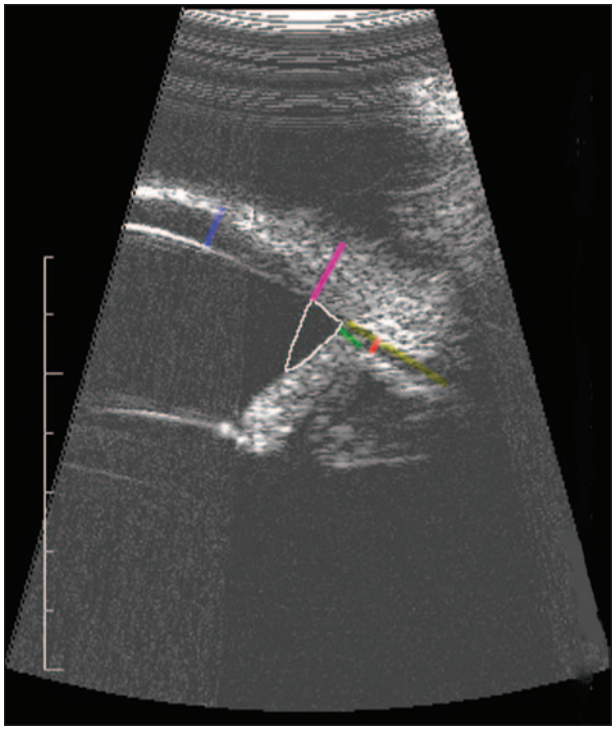

Figure 1—High-resolution ultrasound image of an eye of a clinically normal dog. Notice the following measurements: angle recess area (white), scleral thickness (pink), corneal thickness (blue), iris thickness (green), maximum length of ciliary cleft (yellow), and maximum width of ciliary cleft (red). Scale bar to left indicates millimeters.

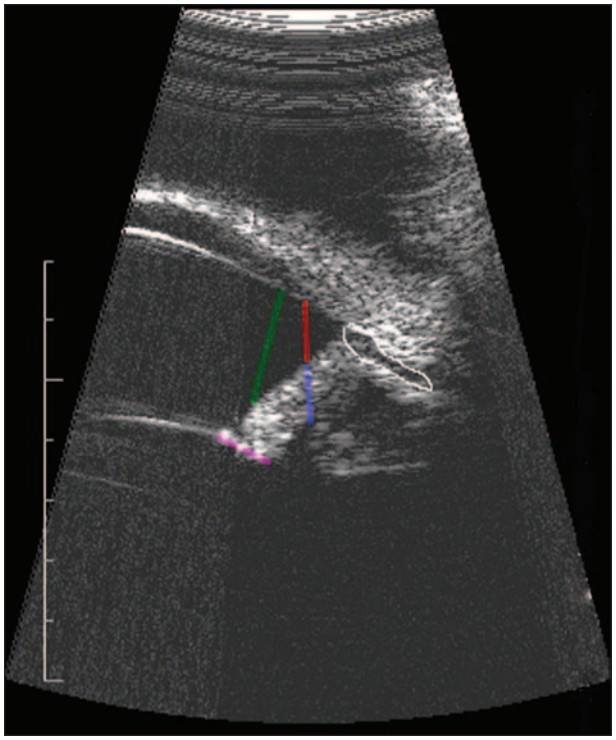

Figure 2—High-resolution ultrasound image of an eye of a clinically normal dog. Notice the following measurements: area of ciliary cleft (white), trabecular meshwork-iris distance (red), iridociliary process distance (red-blue); angle opening distance at 500 μm (green); and iris-lens contact (pink). Scale bar to left indicates millimeters.

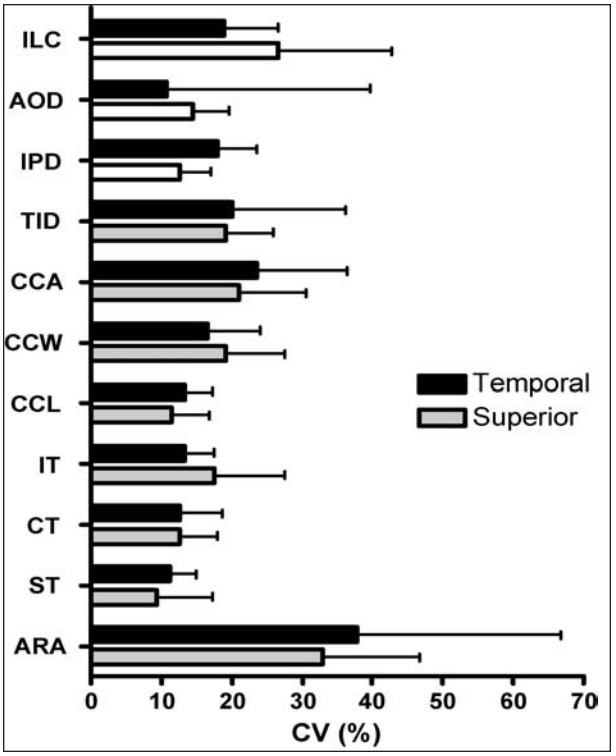

Figure 3—Mean ± SD coefficients of variation of anterior segment structures measured at the superior and temporal locations. ILC = Iris-lens contact. AOD = Angle opening distance at 500 μm. IPD = Iridociliary process distance. TID = Trabecular meshwork-iris distance. CCA = Area of ciliary cleft. CCW = Maximum width of ciliary cleft. CCL = Maximum length of ciliary cleft. IT = Peripheral iris thickness. CT = Corneal thickness. ST= Scleral thickness. ARA = Angle recess area. CV = Coefficient of variation.

ferences were not large between the superior and temporal positions (**Figure 3**). The ARA had the highest CV, whereas ST had the lowest CV of the parameters measured in the superior and temporal positions.

Discussion

Intraobserver reliability was generally good for most parameters measured. A variety of factors contributed to the lesser intraobserver reliability of ARA, CCW, and CCA found in our study. First, these parameters are somewhat ill-defined structures without well-demarcated borders. Furthermore, measurements of ARA, CCW, and CCA involved multiple steps with multiple interpretations of anatomic landmarks that further compounded measuring accuracy. Previous reports[7,9,10] suggest that measurement variability is less on structures that have clearly defined borders; therefore, the use of clearly defined landmarks and structures will increase the accuracy of measurements. Measurements of ST, CT, and iris-lens contact, which have well-defined structures or relationships, had high intraobserver reliability, demonstrating the importance of well-defined structures with few steps in obtaining accurate measurements.

Many measurements in our study were based on the end of Descemet's membrane, which is usually a clearly defined structure in high-resolution images of the canine eye. As a result, lower CVs were achieved for the TID and angle opening distance at 500 μm for

each observer in our study than for similar measurements in previously reported studies[6,7] in humans. Consideration should be given to the anterior segment structure being measured so that structures important to the measurement, such as Descemet's membrane, are clearly imaged.

Despite low CVs, significant differences in measurements of the same image for many of the parameters were found between observers. These discrepancies are likely the result of slight differences in anatomic landmark selection and from idiosyncratic ways of obtaining measurements by each observer. For example, an observer may not mark the exact same spot as the end of Descemet's membrane as another observer, which would lead to substantial differences in measurements between observers because many measurements are based on the end of Descemet's membrane. Also, when structures are poorly defined, each observer will have a slightly different interpretation of the borders of the structure, which likely accounts for the differences in ARA, CCW, and CCA measurements between observers. The low CV for measurements obtained by both observers, however, indicates that results are consistent for a given observer.

Differences between measurements of the 2 observers were not significant for 6 parameters in our study. Considering that the power of our study was 0.8, the mean detectable difference was 10% between measurements obtained by the 2 observers for these 6 anterior segment structures, a finding that should be taken into consideration if multiple observers measure images for quantitative studies.

Image reproducibility was not as consistent as intraobserver reliability for measurements. The CVs for measurements obtained by an observer from multiple images of an eye were higher than those obtained repeatedly by the same observer from a single image, indicating that variability exists between images. In general, CVs for measurements obtained from the superior position were lower than those obtained from the temporal position. For research studies in which repeated ultrasound images are being obtained, these results suggest that the superior position is preferable to obtain the most reproducible results.

Technical reasons for image variation include differences in probe position, probe movement, ocular movement (both intra- and extraocular), and patient movement. Every effort should be made to decrease these variables when attempting to produce images for use in quantitative high-resolution ultrasonographic studies. Furthermore, to obtain the clearest definition of a structure, the transducer needs to be perpendicular to the structure. Unfortunately, it is not possible for all iridocorneal angle structures to be perpendicular to the transducer, which will always result in some compromise to image clarity; therefore, measurement precision for angle structures as structures and borders will not be as clearly defined. High CVs are found for iridocorneal angle measurements, such as ARA, CCW, CCA, and TID, which are likely a result of variations in image clarity.

Anatomic variation is likely another source of image variation. Distinguishing variability as a result of technique from an inherent variation in the tissue itself is difficult. High-resolution ultrasonography results in thin slices of imaged tissue, and repositioning the probe to image the same exact area is difficult. Past work demonstrates wide variation between quadrants in anatomic relationships, such as the distance from the limbus to the pars plana,[11] which suggests that variation likely occurs within quadrants as well.

In general, CVs for measurements in our study were somewhat higher than those found in previously published ultrasound biomicroscopy studies[5-7] in humans. Examination of dogs as opposed to humans is a likely cause of higher CVs for measurements in our study. Generally, human patients are more cooperative and can be directed to fixate on a specific spot during their ultrasonographic examination. Although slight movements no doubt persist, these movements are much less than the ocular movements of a conscious, restrained dog that cannot be instructed to fixate. Another cause of higher CV for measurements in our study is the higher resolution of ultrasound biomicroscopy, which generally uses a 50- to 60-MHz probe, compared with HRUS, which uses a 20-MHz probe. Higher-resolution images from an ultrasound biomicroscope will provide more clearly defined borders to the structures being measured, which, as discussed, will result in less variability in measurements. Ultrasound biomcroscopes, however, usually have fixed handpieces, which makes patient positioning crucial to obtaining an ultrasound biomicroscopic image. As a result, ultrasound biomicroscopy requires sedation or general anesthesia in canine patients.[12] High-resolution ultrasound images, however, can be easily obtained in awake dogs with a handheld ultrasound probe.[1] This ease of use makes HRUS much more practical in a clinical and research setting, as examinations can be performed minutes to hours apart without artifacts in pupil size and accommodation induced by sedation or anesthesia. Eliminating the need for sedation or anesthesia also decreases any associated complications and decreases the expense associated with the examination.

Although variation in measuring and obtaining images exists with HRUS, this does not mean that quantitative studies cannot be performed with this instrument. A large CV will result in a wide distribution around the mean; however, statistical differences may still be measured, provided the difference between the 2 means is sufficiently large. When performing quantitative studies, every effort should be made to design measurements that involve clearly defined, yet relevent structures, and images obtained should have the greatest clarity achievable. Anatomic landmarks used in measurement should be considered while obtaining the images so that every effort can be made to have those landmarks be as clearly identifiable as possible in the obtained images. One operator should obtain the images whenever possible, and only a single observer should measure the images.

a. I^3 B-scan, Innovative Imaging Inc, Sacramento, Calif.
b. Bausch & Lomb Pharmaceuticals, Tampa, Fla.
c. NIH Image, version 1.62 Bethesda, Md. Available at: rsb.info.nih.gov/nih-image. Accessed 2002–2004.

References

1. Bentley E, Miller PE, Diehl KA. Use of high-resolution ultrasound as a diagnostic tool in veterinary ophthalmology. *J Am Vet Med Assoc* 2003;223:1617–1622.

2. Pavlin CS, Foster FJ. Ultrasound biomicroscopy: high-frequency ultrasound imaging of the eye at microscopic resolution. *Radiol Clin North Am* 1998;36:1047–1058.

3. Foster FJ, Pavlin CS. Advances in ultrasound biomicroscopy. *Ultrasound Med Biol* 2000;26:1–27.

4. Ishikawa H, Liebmann J, Ritch R. Quantitative assessment of the anterior segment using ultrasound biomicroscopy. *Curr Opin Ophthalmol* 2000;11:133–139.

5. Urbak SF. Ultrasound biomicroscopy. I. Precision of measurements. *Acta Ophthalmol Scand* 1998;76:447–455.

6. Urbak SF, Pedersen JK, Thorsen TT. Ultrasound biomicroscopy. II. Intraobserver and interobserver reproducibility of measurements. *Acta Ophthalmol Scand* 1998;76:546–549.

7. Tello C, Liebmann J, Potash SD, et al. Measurement of ultrasound biomicroscopy images: intraobserver and interobserver reliability. *Invest Opthalmol Vis Sci* 1994;35:3549–3552.

8. Balidis MO, Bunce C, Boboridis K, et al. Intraobserver and interobserver reliability of the R/D score for evaluation of iris configuration by ultrasound biomicroscopy, in patients with pigment dispersion syndrome. *Eye* 2002;16:722–726.

9. Ishikawa H, Schuman M. Anterior segment imaging: ultrasound biomicroscopy. *Ophthalmol Clin North Am* 2004;17:7–20.

10. Urbak SF. Ultrasound biomicroscopy. III. Accuracy and agreement of measurements. *Acta Ophthalmol Scand* 1999;77:293–297.

11. Smith P, Pennea L, Mackay E, et al. Identification of sclerotomy sites for posterior segment surgery in the dog. *Vet Comp Ophthalmol* 1997;7:180–191.

12. Gibson TE, Roberts SM, Severin GA, et al. Comparison of gonioscopy and ultrasound biomicroscopy for evaluating the iridocorneal angle in dogs. *J Am Vet Med Assoc* 1998;213:635–638.