

# Radiographic Heart Size and Its Rate of Increase as Tests for Onset of Congestive Heart Failure in Cavalier King Charles Spaniels with Mitral Valve Regurgitation

P.F. Lord, K. Hansson, C. Carnabuci, C. Kwart, and J Haggström

**Background:** In canine mitral regurgitation (MR) the rate of heart enlargement increases in the last year before congestive heart failure (CHF). Measurement of heart size and its rate of increase may be useful tests for CHF in MR.

**Objectives:** To determine the value of vertebral heart scale (VHS) and its rate of increase ( $\Delta$ VHS units/month) for diagnosing the presence and predicting the onset of CHF.

**Animals:** Longitudinal study of 94 Cavalier King Charles Spaniels (CKCS).

**Methods:** VHS was measured at intervals before CHF.  $\Delta$ VHS/month was calculated from sequential pairs of VHS measurements and the interval between them. Diagnostic accuracy and utility were determined by the areas under receiver operating characteristic plots (AUROC), and likelihood ratios (LR).

**Results:** AUROC for VHS at the onset of CHF was 0.93 (95% CI, 0.96–0.90), to predict CHF 1–12 months before CHF was 0.74 (95% CI, 0.81–0.66), and for  $\Delta$ VHS/month at CHF was 0.98 (95% CI, 0.99–0.96). Interval LRs and their cutoff values for CHF were for VHS: 13 (95% CI, 20–7.3) at  $\geq 12.7$ ; 1.2 (95% CI, 2.0–0.68) between 12.7 and 12.0; 0.04 (95% CI, 0.18–0.01) at  $\leq 12.0$ , and for  $\Delta$ VHS/month: 15 (95% CI, 30–7.7) at  $\geq 0.08$ ; 0.72 (95% CI, 2.0–0.25) between 0.08 and 0.06; and 0.05 (95% CI, 0.13–0.02) at  $\leq 0.06$ .

**Conclusions and Clinical Importance:** Under the conditions of this study, VHS and particularly  $\Delta$ VHS/month are useful measurements for detecting onset of CHF in CKCS with MR.

**Key words:** Diagnosis; Likelihood ratio; Vertebral heart scale.

Mitral valve regurgitation (MR) caused by myxomatous mitral valve degeneration (MMVD) is a common disease of small dogs, causing morbidity and mortality attributable to congestive heart failure (CHF). Many dogs with murmurs never develop CHF or do so only years after a MR murmur is heard.<sup>1,2</sup> The diagnosis of CHF is not always straightforward. The clinical and radiographic<sup>3</sup> signs may not be definitive and comorbidity of respiratory diseases may decrease the accuracy of diagnosis. It has not been possible to predict if or when CHF will develop.

In a recent longitudinal study of 24 CKCS, values of VHS and its rate of increase ( $\Delta$ VHS/month) at CHF were well separated from values before the onset of CHF, but the number of dogs was insufficient for a statistically reliable evaluation of their use in diagnosing the onset of CHF.<sup>4</sup> Yearly radiographic or echocardiographic monitoring recently was recommended to evaluate for rapidly increasing size of the left atrium as evidence of impending CHF,<sup>5</sup> indicating awareness

## Abbreviations:

AUROC	area under the receiver operating characteristic curve or plot
CKCS	Cavalier King Charles Spaniels
CHF	congestive heart failure
CR	coefficient of repeatability
LR	likelihood ratio
MMVD	myxomatous mitral valve degeneration
MR	mitral valve regurgitation
ROC	receiver operating characteristic
VHS	vertebral heart scale
$\Delta$ VHS/month	increase of VHS units per month

of the clinical relevance of the rate of increase of heart chamber size.

The receiver operating characteristic curve (ROC) is the best single statistical method to evaluate a diagnostic test, as it incorporates all levels of specificity and sensitivity and is independent of the prevalence of the disease in the population.<sup>6–9</sup> Although the area under the ROC (AUROC) curve or plot is the best summary of the accuracy of a diagnostic test,<sup>6–10</sup> and AUROC can be used to compare tests done on the same or similar patient groups, it cannot directly be used to determine the effect of a test result on the individual patient.<sup>11</sup> Vertebral heart scale (VHS) as a measurement of heart size discriminated between CHF and noncardiac causes of coughing in dogs with MR, with an AUROC of 0.92.<sup>12</sup> A value of 1.0 is a perfect test and 0.5 is no better than chance. A diagnostic test should be done when the pretest probability of the condition being present is between thresholds for treating the disease in question or ruling out the condition.<sup>10,13</sup> A good test substantially changes the probability of the disease being present or absent compared

*From the Department of Clinical Sciences (Lord, Hansson, Carnabuci, Haggström) and the Department of Anatomy and Physiology (Kwart), Faculty of Veterinary Medicine and Animal Science, Swedish University of Agricultural Sciences, Uppsala, Sweden. The work was done at the Department of Clinical Sciences, Faculty of Veterinary Medicine and Animal Husbandry, Swedish University of Agricultural Sciences. No grant support. Not presented at a meeting.*

*Corresponding author: P. Lord, Swedish University of Agricultural Sciences, P.O. Box 7054, SE-750 07 Uppsala, Sweden; e-mail: peter.lord@slu.se*

*Submitted December 16, 2010; Revised June 28, 2011; Accepted August 1, 2011.*

*Copyright © 2011 by the American College of Veterinary Internal Medicine*

*10.1111/j.1939-1676.2011.00792.x*

to the probability before the test. The quantification of this change is the likelihood ratio (LR).<sup>6-8,10,11,13-15</sup> The LR is the factor by which the pretest probability of a disease being present or absent in a patient is changed by a test result to give a post-test probability and is the ratio of the probability that a specific test result is obtained in a patient with or without the condition divided by the probability of obtaining the same test result in a patient without or with the condition. The use of LRs to calculate post-test probability is an application of Bayes' theorem.<sup>10,15,16</sup>

The hypothesis of this study was that the VHS and its rate of increase could discriminate between the onset of CHF and pre-CHF MR while monitoring dogs yearly with radiography, and that VHS could predict if a dog would develop CHF within a year of the examination. The objectives were to determine the accuracy of VHS and its rate of increase, using AUROC plots and LRs. An additional aim was to determine if sequential measurements of VHS had less variability than random measurements.

## Materials and Methods

VHS was measured on left lateral radiographs of 94 (56 males, 38 females) Cavalier King Charles Spaniels (CKCS) examined sequentially at time intervals of 12 months from 1 to 5 times before the dogs developed CHF and at the time of presentation for clinical signs of CHF, before any treatment for CHF was started. The examinations were made as part of a previous study.<sup>2</sup> All dogs in the study that developed CHF and were radiographed at the onset of CHF and at scheduled intervals before treatment were included. The radiographs were taken at the 12 practices that participated in the study. Quality of radiographs and criteria for pulmonary edema were reviewed at annual meetings of the participants. In addition, at the end of the study all radiographs were reviewed for quality of exposure and positioning by 1 radiologist (KH), who found that all were acceptable for measurement of VHS and diagnosis of pulmonary edema. Dogs with comorbidities were excluded. A diagnosis of the onset of CHF required (1) owners' complaints of dyspnea, cough, exercise intolerance, and nocturnal restlessness, (2) physical examination confirming dyspnea, tachypnea, and loss of respiratory sinus arrhythmia on examination, and (3) radiographic signs of pulmonary edema with cardiac and left atrial enlargement, and exclusion of other causes of coughing or dyspnea. Accurate diagnosis of pulmonary edema was aided by use of the same exposure at each examination and comparison of lung opacity with that of radiographs made previously. Pulmonary edema was a requirement and confirmed by blinded review of all radiographs by 1 investigator (KH) at the end of the study. As further evidence of the correct diagnosis of the onset of congestive heart failure, all dogs either responded to treatment for pulmonary edema after the diagnosis of the onset of CHF, or CHF was confirmed at necropsy if the dog was euthanized. Neither subjectively evaluated degree of heart enlargement nor VHS was the criterion for the reference standard of CHF as the end point. Forty-three dogs were being treated with 0.25–0.50 mg/kg (mean  $\pm$  SD,  $0.37 \pm 0.08$  mg/kg PO daily) enalapril for the clinical trial for which they had been recruited. No other treatment was given until after the radiographs at onset of CHF had been made.

VHS was measured by a slight modification<sup>17</sup> of the original method,<sup>18</sup> to improve the precision of the anatomical points for measurement. The starting point for the short axis was the middle of the caudal vena cava where it joined the heart. The long

axis was measured from the ventral border of the left cranial lobe bronchus seen in cross-section (end-on). When it was not clear which was the left bronchus, the more cranial bronchus was chosen. All VHS measurements of each dog were made by 1 investigator (PL) 3 times at 1 occasion, and the result averaged. For all but the 1st radiograph of each sequence of radiographs, the anatomic points used on the preceding radiograph for measurement of the axes provided a guide to consistent location of the anatomic points for each following radiograph.

To determine if accuracy of measuring improved with experience (i.e., a learning effect), an experienced (PL) and inexperienced (CC) observer measured VHS 4 times each on sets of sequential radiographs of 6 of the dogs in this study (a total of 78 radiographs), with an interval of at least 4 weeks between measuring the sets. The inexperienced person was given written instructions only, as described.<sup>17</sup> The radiographs were then randomized and the same observers measured them twice, with an interval of at least 4 weeks between measurements. For each observer, the coefficients of repeatability (CR)<sup>19</sup> of sequential pairs of measurements (1–2, 2–3, 3–4) and the pairs of random measurements were calculated. The interobserver CR of the 2 observers was calculated by comparing all pairs of sequential measurements made by the two observers. The intraclass correlation coefficient (ICC) of the sequential measurements was calculated for each observer.

Rate of increase of VHS was calculated as  $\Delta$ VHS/month by dividing the difference between every consecutive pair of VHS values at each interval between measurements by the number of months of the interval. In serial measurements of a group of patients, the rate of change of a measurement may be calculated as a summary statistic from each patient's data, and analyzed as though it were raw data.<sup>20</sup>

The possible effect of treatment with enalapril on VHS at CHF and  $\Delta$ VHS/month at the last time interval was evaluated by the Mann–Whitney *U*-test for non-normally distributed independent samples with statistical significance set at  $P < .05$ . Because males had a statistically insignificant tendency to a more rapid progression to CHF than did females,<sup>2</sup> we compared  $\Delta$ VHS/month for males and females using the same test. Because treatment with enalapril had no effect on VHS ( $P = .19$ ) or  $\Delta$ VHS/month at CHF ( $P = .77$ ) and sex did not affect  $\Delta$ VHS/month at CHF ( $P = .70$ ), all dogs were treated as 1 group in all subsequent analyses.

The following data were used to produce results for 5 tests:

1. VHS at onset of CHF (time 0) compared with all pre-CHF VHS (at times 1–5).
2. VHS at onset of CHF compared with VHS measured at the last scheduled examinations before CHF (time 1), all of which were at  $\leq 1$  year before CHF. This was likely to be a more difficult discrimination than Test 1, because the mean VHS of the negative group at time 1 would be higher than the mean VHS of the negative group in Test 1, in which the negative group encompassed durations up to 5 years before CHF. Test 2 was done to see how much VHS lost accuracy when the dogs in the negative group were near onset of CHF and their hearts were larger than in Test 1.
3. VHS at time 1 compared with the VHS at times 2–5. Because time 1 was designated as a positive result, this test predicted onset of CHF within the interval time 1 (less than 1 year to onset of CHF), compared to all dogs at times 2–5.
4.  $\Delta$ VHS/month at onset of CHF compared with all pre-CHF  $\Delta$ VHS/month.
5.  $\Delta$ VHS/month at onset of CHF compared with  $\Delta$ VHS/month at interval 1, the second last interval before CHF. This is a more difficult test, because only dogs within 1 year of

onset of CHF were included in the negative group and  $\Delta$ VHS/month would be closer to the values at CHF than when all dogs without CHF were included. Test 5 was done to see how much  $\Delta$ VHS/month lost accuracy as the dogs in the negative group approached onset of CHF and  $\Delta$ VHS/month would be expected to be higher than in Test 4.

**Statistical Methods**

ROCs were plotted for tests 1 and 3–5 using a method that accounts for the repeated measurement of VHS and  $\Delta$ VHS/month of the same dog over time to monitor for the occurrence of the event of CHF, and a variable number of measurements per individual, with only 1 possible observation with true positive disease status (CHF).<sup>21</sup> This method decreases the variance of the test compared with using only 1 time point. Test 2 had only positive values at time 0 and only negative values at time 1, so the method of DeLong<sup>22</sup> was used instead. The 95% confidence intervals (CI) of the AUROC were calculated by the jackknife method.<sup>23</sup>

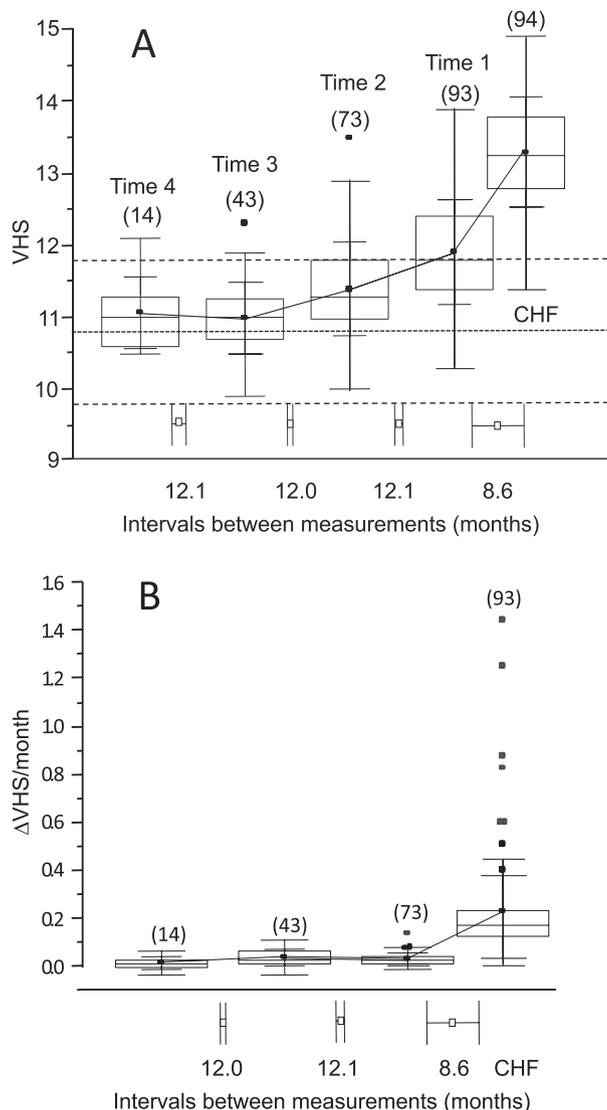
LRs can be derived by dichotomizing the results at a chosen cutoff point, usually at the maximum Youden index ( $J = \text{Sensitivity} + \text{Specificity} - 1$ ) where the vertical distance between the ROC curve and the diagonal or chance line connecting (0,0) and (1,1) is maximal,<sup>7,9</sup> or the point closest to the top left corner of the graph. Unless the test is very accurate, dichotomizing continuous results loses information,<sup>10,11,24–26</sup> as values near the cutoff value are as equally weighted as are values near the extremes of the range, which is misleading. The range of continuous values on the ROC plot between clinically useful intervals can be used to calculate interval LRs,<sup>24,25,27</sup> which separate the values at the steepest and flattest parts of the plot that give the most reliable information (interval 1 and 3) from the intermediate interval (interval 2) which comprises inconclusive values.

Dichotomous and interval LRs and their 95% CI at cutoff values based on the shape of the ROC plots of VHS and  $\Delta$ VHS/month were calculated for tests 1, 3, and 4. The optimal cutoff value for the dichotomized tests was chosen for tests 3 and 4 at the highest Youden index. For test 1 we used a value that equalized sensitivity and specificity and still was close to the maximum Youden index, which gave a value (12.8) that gave high specificity at the expense of sensitivity. By using the range of continuous values on the ROC plot, the cutoff points at clinically useful intervals were used to calculate interval LRs. Two natural breaks in the plot were selected to make 3 intervals: Interval 1: (0,x), Interval 2: (x,y) and Interval 3: (y,I).<sup>24,25,28</sup> LRs for each interval and their 95% CIs were calculated.<sup>27</sup>

ROCs and AUROCs with their 95% CIs were calculated using a macro<sup>21</sup> for SAS.<sup>a</sup> All other statistics were calculated with Medcalc for Windows 11.4©.<sup>b</sup>

**Results**

The values of VHS at the times of examination and  $\Delta$ VHS/month at each of the time intervals are shown as box plots (Fig 1). The time intervals were consistently close to the scheduled 1 year except for the last interval before the onset of CHF. This was shorter than 1 year and more variable because the dogs were presented with CHF at variable times before the next scheduled examination. VHS steadily increased until the last measurement when it was highest.  $\Delta$ VHS/month did not increase until the last interval, which ended with the onset of CHF.



**Fig 1.** Box and whisker plots of vertebral heart score (VHS) of 94 dogs (A) and rate of change of VHS ( $\Delta$ VHS/month) in 93 dogs (B) at each measurement before the onset of congestive heart failure. The number near the box is the number of dogs. Small black squares are outliers. The central box represents the values from the lower to upper quartile (25–75 percentile). The middle line represents the median. The vertical line extends from the minimum to the maximum value, excluding outliers. The broken lines in (A) are the mean  $\pm$  2 SD of 10 normal Cavalier King Charles Spaniels.<sup>17</sup> The median VHS for all values before onset of CHF was 11.5 (95% confidence interval CI, 11.3–11.6). The median VHS at CHF was 13.3 (95% CI, 13.1–13.4). The median  $\Delta$ VHS/month of all values before CHF was 0.03 (95% CI, 0.02–0.03). At CHF the median was 0.17 (95% CI, 0.15–0.19). The 5th time and 4th interval were omitted as they had only 4 dogs.

The ROC plots for tests 1, 3, and 4 are shown in Figure 2. The AUROC for all of the tests and their related statistics are listed in Table 1. Differences in AUROC between tests 1 and 2, and 4 and 5 were not significant ( $P > 0.05$ ).

The LRs, their cutoff values, and CIs for tests 1, 3, and 4 are shown in Table 2. Moving indeterminate

values to Interval 2 ( $x,y$ ) did not increase the 95% CIs of the useful interval LRs ( $0,x$ ) and ( $y,1$ ) (Intervals 1 and 3). VHS/month had a much smaller intermediate range (Interval 2) than did VHS (Fig 2).

Sequential paired comparisons for 2 observers showed no decrease from the 1st to the 4th examination (Table 3), indicating that no learning effect was found, and the ICC could be applied to compare the observers. The inexperienced observer's results were as good as those of the experienced observer. ICC for sequential measurements of VHS for both the experienced and inexperienced observers were close, with overlapping CI (Table 3).

## Discussion

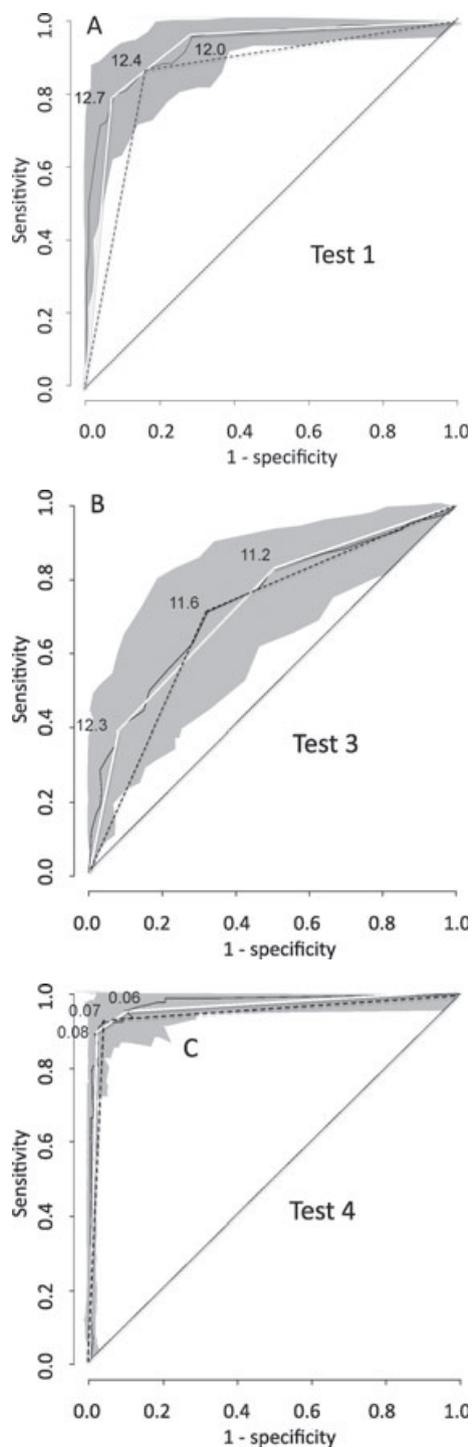
The value of the AUROC as a criterion of the accuracy of a test has been stated as being low, 0.5–0.7; moderate, 0.7–0.9; or high, >0.9.<sup>13</sup> Thus, tests 1, 2, 4, and 5 were highly accurate. The AUROCs of tests 2 and 5 showed that VHS and  $\Delta$ VHS/month still were highly accurate tests when the difficulty of discriminating between pre-CHF and onset of CHF increased, when the negative dogs were within a year of onset of CHF. However, their 95% CIs were greater because of more overlapping of results and fewer samples in the negative groups. These 2 tests cannot be used on patients because the time to CHF cannot be known. Test 3 was moderately accurate, but the CIs were wide (AUROC = 0.74, 95% CI, 0.81–0.66).

### *Effect of Sequential Measurement on VHS Variability*

The results in Table 3 suggest that variability of VHS is not improved by experience. The greater mean difference of the pairs of random measurements indicates that sequential measurements slightly decreased variability. Variability of VHS and  $\Delta$ VHS/month should be investigated using more observers.

### *Likelihood Ratios*

Because the scaling of LRs is logarithmic, LRs of >10 or <0.1 make a large and often conclusive change in pretest probability and the diagnosis, LRs between 5 and 10, and 0.1 and 0.2, cause moderate changes in the pretest probability, LRs between 2 to 5 and 0.5 and 0.2 generate small changes, and those between 0.2 and 0.5 have very little effect on pretest probability.<sup>13</sup> In this study, VHS dichotomized at 12.4 units had a LR+ of 7.3 and a LR– of 0.15 units (Test 1, Table 2). Using interval LRs improved the LRs of the tests to 13 and 0.04, because indeterminate values were moved to Interval 2. For Test 3, the wide CIs caused by fewer observations in the 3 intervals brought the limits of LRs of Intervals 1 and 3 too close to 1.0 (Table 2) for them to have a useful effect on diagnosis, particularly considering the difficulty of determining the pretest probability of a dog with an enlarged heart but without clinical signs developing CHF within a year.



**Fig 2.** ROC plots for tests 1, (A), 3, (B), and 4, (C). The number on the plot closest to the top left corner of the graphs (1,0), is the cutoff value for a dichotomized test. The broken lines are the intervals from this cutoff value to the lower left (0,0) and upper right (1,1) coordinates. The 2 other numbers are cutoff values selected at the changes in slope of the plot (white lines). The area under the diagonal dashed line from (0,0) to (1,1) is 0.5 of the total area, and the line indicates a test no better than chance. The gray zone shows the 95% CIs of the ROC plots.

**Table 1.** Accuracy of the 5 tests expressed as area under the receiver operating characteristic plots.

Test Number	Method	Sample Size	AUROC Plot	95% CI
1	VHS at CHF versus all pre-CHF	94+/227-	0.93	0.96-0.90
2	VHS versus <1 year to CHF	94+/93-	0.91	0.94-0.85
3	VHS <1 year to CHF versus all >1 year to CHF	93+/134-	0.74	0.81-0.66
4	$\Delta$ VHS/month at CHF versus all pre-CHF	93+/134-	0.98	0.99-0.96
5	$\Delta$ VHS/month at CHF versus <1 year to CHF	93+/73-	0.96	1.00-0.93

AUROC, area under the receiver operating characteristic plot as fraction of total possible area; VHS, vertebral heart scale; CHF, congestive heart failure; CI, confidence interval;  $\Delta$ VHS/month, rate of increase of VHS per month.

**Table 2.** Likelihood ratios of tests 1, 3, 4 at the cutoff points shown on Figure 2.

Test	Cutoff Value	Number Positive	Number Negative	LR	95% CI	
1. VHS at CHF dichotomized	>12.4	82	27	7.3 (LR+)	10.5-5.1	
	<12.4	12	200	0.15 (LR-)	0.25-0.09	
1. VHS at CHF, Interval 1	$\geq$ 12.7	75	18	13	20-7.3	
	Interval 2	12.7-12.0	16	33	1.2	2.0-0.68
	Interval 3	$\leq$ 12.0	3	176	0.04	0.18-0.01
3. VHS <1 year before CHF, dichotomized	>11.6	57	31	2.7 (LR+)	3.8-1.9	
	<11.6	36	103	0.50 (LR-)	0.66-0.38	
3. VHS < 1 year before CHF, Interval 1	$\geq$ 12.3	27	8	4.9	10-2.3	
	Interval 2	11.2-12.3	49	49	1.4	1.9-1.1
	Interval 3	$\leq$ 11.2	17	77	0.32	0.5-0.20
4. $\Delta$ VHS/month at CHF, dichotomized	>0.07	87	10	13 (LR+)	23-6.9	
	<0.07	6	124	0.07 (LR-)	0.15-0.03	
4. $\Delta$ VHS/month at CHF, Interval 1	$\geq$ 0.08	84	8	15	30-7.7	
	Interval 2	0.06-0.08	5	10	0.72	2.0-0.25
	Interval 3	$\leq$ 0.06	4	116	0.05	0.13-0.02

VHS, vertebral heart scale; LR, likelihood ratio; CI, confidence interval.

LR+ and LR- are applied only to dichotomized data. The positive and negative numbers were used in a 2 x 2 (dichotomized) and 2 x 3 (interval) matrix to calculate LRs. The precision of the LRs varies according to the magnitude because the scale is logarithmic.

**Table 3.** Statistics of comparison of sequential and random measurements of vertebral heart scale by an experienced and an inexperienced observer.

Coefficient of Repeatability	Sequential 1-2	Sequential 2-3	Sequential 3-4	Random	Inexperienced versus Experienced Observer (sequential measurements)	
					ICC for Sequential Measurements (single measures)	
Inexperienced observer	0.66	0.65	0.72	0.76	0.933 (95% CI, 0.888-0.959)	
Experienced observer	0.60	0.78	0.58	0.69	0.90	0.943 (95% CI, 0.92-0.96)

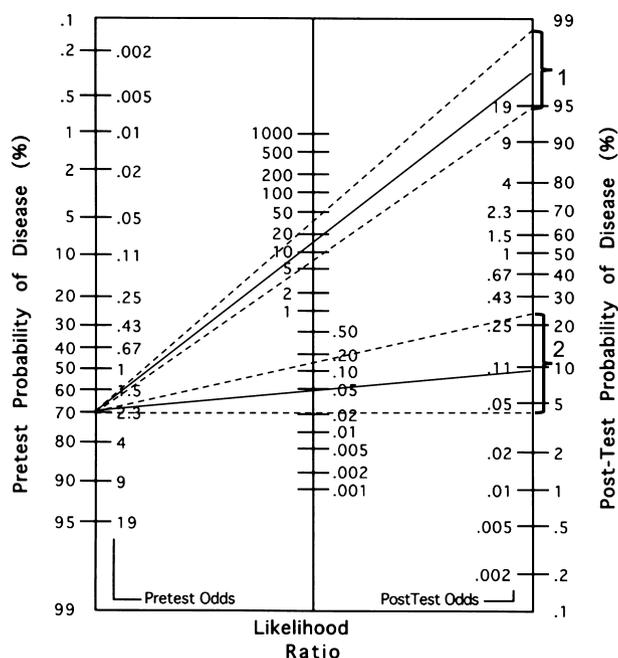
ICC, intraclass correlation coefficient.

The coefficient of repeatability is for comparisons between sequential paired measurements from 1 to 4.

$\Delta$ VHS/month values (Test 4, Table 2), when dichotomized at 0.07, were 13 (LR+) and 0.07 (LR-), and improved slightly by using intervals of  $\geq$ 0.08 and  $\leq$ 0.06 to 15 and 0.05, with minimal increase in the CIs. Because this test was already excellent, only a small interval of indeterminate values ( $>$ 0.06 and  $<$ 0.08 VHS units/month) needed to be removed. In practice, 2 decimal points is the limit of accuracy, so 0.07 is the indeterminate value. Values of  $\Delta$ VHS/month  $\geq$ 0.08 or  $\leq$ 0.06 have a large effect on pretest probability of CHF.

In tests 1, 3, and 4, Intervals 1 and 3 were farther from the neutral value of 1.0 than that of the dichoto-

mized LR+ and LR-, evidence of the superiority of interval LRs over dichotomized LRs. Despite fewer observations, the removal of indeterminate values into Interval 2 did not lower the accuracy of the tests for the onset of CHF. As expected from the design, the 95% CIs of the LRs of Interval 2 of tests 1, 3, and 4 were in the range considered to have little effect on the pretest probability. Important principles when selecting cutoff values for interval LRs are that the CIs of interval LRs should not overlap substantially, the LRs must be monotonic (in increasing or decreasing order), and that indeterminate interval LRs should have CIs that include 1.0.<sup>11,27</sup> The interval LRs fulfilled these criteria.



**Fig 3.** Example of use of and likelihood ratios (LR) and a nomogram<sup>31</sup> with VHS velocity to determine 2 post-test probabilities from a pretest probability of 70%. In Example 1 (bracket 1), when the VHS velocity is 0.09 units/month or greater (Interval 1), the LR is 15 (95% confidence interval CI, 30–8) (Table 2). By drawing a straight line (solid line) from the pretest probability of 70% on the left scale, through the LR of 15 in the middle scale and extending it to the scale on the right, the post-test probability is shown as 97% (95% CI, 99–95%). Dashed lines are 95% CIs. In Example 2 (bracket 2), in which VHS velocity is 0.04  $\Delta$ VHS/month (Interval 3), the LR is 0.05 (95% CI, 0.13–0.02) (Table 2). The post-test probability is now 10% (95% CI, 25–4%). Note that the scales are logarithmic. Copyright © 1975 Massachusetts Medical Society. All rights reserved.

We used LR to evaluate the tests rather than the more common specificity and sensitivity at selected cutoff values because LR incorporates both sensitivity and specificity. Sensitivity and specificity evaluate the test given the patient's status, whereas the clinician needs to know the patient's status given a test result.<sup>11,29</sup> Predictive values are highly dependent on the prevalence of the condition, and prevalence is not the same in the test population as in the real population. LR is independent of prevalence, this being estimated separately as pretest probability in an individual patient.<sup>11</sup> LR refines the clinical judgment when the pretest probability is in the uncertain range where the diagnosis is not excluded and yet is not certain enough to treat the patient.<sup>30</sup> Subjective ordinal terms used clinically such as low, intermediate, and high probability can be converted to ranges of numerical probabilities.<sup>6,29,30</sup> Finally, interval LR improve accuracy by removing indeterminate values.<sup>6,10,24–27</sup>

Differentiating VHS by time has advantages for use in breeds other than CKCS. Differentiation removes

systematic between-dog variations in VHS caused by breed and anatomical differences, including stable disk degeneration shortening the thoracic spine, and inter-reader variations in choosing anatomical points for measurement. These are likely even when the criteria are clearly stated.<sup>17</sup> Although  $\Delta$ VHS/month is likely to remain accurate when used with other breeds, this hypothesis has to be evaluated. The rate of change of a dimension is its velocity, regardless of the units. Hence, we propose the term “VHS velocity” rather than “rate of change” or “rate of increase.”

### Example of Use of LR in a Hypothetical Patient

A 10-year-old CKCS has been monitored yearly for 3 years, according to the recommendation of the *ACVIM Consensus Statement* for MR,<sup>5</sup> because of a MR murmur and enlarged heart on radiographs. Seven months after the last scheduled examination it is radiographed for clinical signs compatible with CHF. The veterinarian estimates that the dog's pretest probability of CHF being present is 70%, because the clinical signs were not specific and the radiographs, although ruling out radiographically apparent comorbidities, were inconclusive for pulmonary edema. This value represents “probably has CHF” on an ordinal or categorical scale.  $\Delta$ VHS/month is measured from the 2 sets of radiographs 7 months apart.

Bayes' theorem requires the conversion of probabilities to odds and back, because the post-test odds of disease equal the pretest odds times the LR. Instead of these calculations, a simple nomogram<sup>31</sup> is used (Fig 3). In Example 1, a VHS velocity of 0.09 units/month (Interval 1) gives a LR of 15 (95% CI, 30–7.7) (Table 2). The post-test probability using the nomogram is 97% (95% CI, 98–95%). The lower CI limit of 95% probability of CHF is above a reasonable threshold to treat this patient. In Example 2, a VHS velocity of 0.04  $\Delta$ VHS/month (Interval 3) gives a LR of 0.05 (95% CI, 0.13–0.02) (Table 2). The post-test probability is 10% (95% CI, 25–4%). The upper 95% CI limit of 25% should be considered when evaluating the result. CHF probably cannot yet be ruled out. Post-test probability is highly dependent on pretest probability, which is not assessed accurately by physicians.<sup>32,33</sup> The same is likely to be true of veterinarians,<sup>6,14</sup> and is a limitation of all tests except those few that are so accurate that they can rule in or rule out a diagnosis regardless of the pretest probability.<sup>34,35</sup>

### Limitations of the Study

**Selection of Patient Material.** The VHS results are affected by spectrum bias<sup>8,13</sup> in that the spectrum of patients was limited to 1 breed, and breeds vary in their reference ranges of VHS.<sup>36–38</sup> VHS in other breeds is biased by the difference between the normal VHS of the other breed and CKCS. This is a systematic bias and the precision is not likely to be decreased.<sup>39</sup> A correction factor of the difference

between the mean value of normal VHS for the breed in question and CKCS could be evaluated. The left lateral view was used. If the right lateral projection is used a systematic bias of 0.2–0.3 VHS units between the left lateral and right lateral projections<sup>40</sup> must be taken into account. VHS is increased by congenital malformations shortening the vertebrae, and disk degeneration narrowing the disk space. Other causes of dog and breed variability would decrease the accuracy of VHS.

VHS velocity is not affected by between-dog variations, but is increased by progression of disk degeneration during the period of monitoring. This variation may be avoided by always normalizing the heart dimensions of each radiograph with the vertebrae on the first radiograph of the series rather than the vertebrae on the measured radiograph. It is possible, but has not been shown, that breed affects the rate of progression of MMVD and MR before the onset of CHF.

**Determination of the Onset of CHF.** Although we used all available clinical and radiographic criteria, with later review of the radiographs by 1 blinded radiologist, there could have been errors, most likely delayed diagnoses, because clinical signs of CHF had to be verified by radiographic signs of pulmonary edema.<sup>2</sup> If the same criteria for the onset of CHF (both clinical and radiographic signs present) are accepted as an accurate reference standard, VHS velocity would be an accurate surrogate for them.

**Comorbidities.** Although we included a spectrum of dogs with different degrees of MR in the negative groups, comorbidities in dogs with MR murmurs that cause cor pulmonale, such as chronic bronchitis, bronchiectasis, heartworm disease, collapsing trachea or bronchus were excluded. To the degree that right ventricular enlargement of cor pulmonale would increase the short axis for VHS measurement, increasing VHS caused by progressive cor pulmonale with MR present would increase the proportion of false positives. Comorbidities may be evaluated by clinical and radiographic criteria to refine the pretest probability of CHF being the cause of the clinical signs compatible with CHF.

**Measurement Variability.** One experienced observer was used. Interobserver variability of experienced users of VHS was not measured. If the variability of inexperienced observers is greater than that of experienced observers, the tests may be less accurate when applied by an inexperienced observer.<sup>13</sup> However, the variability of an inexperienced observer may not be greater (Table 3).<sup>17</sup> Measurement error (intraobserver variability) of the averaged 3 measurements of VHS was incorporated in the 95% CIs. Systematic bias<sup>39</sup> between observers caused by choice of landmarks will affect accuracy of VHS<sup>17</sup> but not of VHS velocity. The apex is sometimes not clearly defined and the long axis can then only be approximated. Another source of variation in practice could be greater inconsistencies in positioning for radiography than in this study, affecting the measurements. Variability of measurement and

ways to decrease it are areas for further investigation. Even with careful corrections and measurements, neither test is likely to be as accurate in a heterogeneous group of dogs with MR.

Finally, the test has not yet been validated on a 2nd independent group of patients in a prospective cohort study.<sup>13</sup>

## Conclusions

VHS is a useful test for onset of CHF in CKCS not yet treated for CHF. VHS velocity is an accurate test for onset of CHF when monitoring untreated CKCS with MR with yearly examinations and it is likely to be applicable to other breeds. Cutoff values of  $\geq 0.08$  and  $\leq 0.06$  VHS units/month had large effects on pretest probability of onset of CHF. The results warrant further evaluation of these indices in a more heterogeneous group of dogs with MR.

---

## Footnotes

<sup>a</sup>SAS 9.2 Software. SAS Institute Inc, Cary, NC

<sup>b</sup>MedCalc© Software Version 11.4, Mariakerke, Belgium 2009

---

## Acknowledgments

We appreciate the assistance of the following veterinarians, Henrik Pedersen, Anders Eriksson, Anna-Kaisa Järvinen, Anna Tidholm, Karina Bsenko, Erik Ahlgren, Mikael Ilves, Björn Åblad, Torkel Falk, Ellen Bjerkås, Susanne Gundler, Gudrun Wegelund, Eva Adolphson, and Jens Corfitzen, who examined the dogs, and took and provided the radiographs.

Corette Parker, Statistics and Epidemiology Division, RTI International, Research Triangle Park, NC, kindly provided the SAS macro.

Lars Berglund, Uppsala Clinical Research Centre, Faculty of Medicine and Pharmacology, Uppsala University, Sweden, did the statistical analyses using the SAS macro.

Christopher Lamb, Royal Veterinary College, London, provided helpful comments on the manuscript.

## References

1. Borgarelli M, Savarino P, Crosara S, et al. Survival characteristics and prognostic variables of dogs with mitral regurgitation attributable to myxomatous valve disease. *J Vet Intern Med* 2008;22:120–128.
2. Kwart C, Häggström J, Pedersen HD, et al. Efficacy of enalapril for prevention of congestive heart failure in dogs with myxomatous valve disease and asymptomatic mitral regurgitation. *J Vet Intern Med* 2002;16:80–88.
3. Hansson K, Häggström J, Kwart C, et al. Reader performance in radiographic diagnosis of signs of mitral regurgitation in cavalier King Charles Spaniels. *J Small Anim Pract* 2009;50:44–53.

4. Lord P, Hansson K, Kwart C, et al. Rate of change of heart size before congestive heart failure in dogs with mitral regurgitation. *J Small Anim Pract* 2010;51:210–218.
5. Atkins C, Bonagura J, Ettinger S, et al. ACVIM Consensus Statement. Guidelines for the diagnosis and treatment of canine chronic valvular heart disease. *J Vet Intern Med* 2009;23:1142–1150.
6. Gardner IA, Greiner M. Receiver-operating characteristic curves and likelihood ratios: Improvements over traditional methods for the evaluation and application of veterinary clinical pathology tests. *Vet Clin Pathol* 2006;35:8–17.
7. Thrusfield M. Chapter 17. Diagnostic testing. *Veterinary Epidemiology*, 3rd ed. Oxford, UK: Blackwell Science Ltd.; 2005:313–329.
8. Drobatz KJ. Measures of accuracy and performance of diagnostic tests. *J Vet Cardiol* 2009;11 Suppl 1:S33–S40.
9. Akobeng AK. Understanding diagnostic tests 3: Receiver operating characteristic curves. *Acta Paediatr* 2007;96:644–647.
10. Fischer JE, Bachmann LM, Jaeschke R. A readers' guide to the interpretation of diagnostic test properties: Clinical example of sepsis. *Intensive Care Med* 2003;29:1043–1051.
11. Gallagher EJ. Evidence-based emergency medicine/editorial. The problem with sensitivity and specificity. *Ann Emerg Med* 2003;42:298–303.
12. Guglielmini C, Diana A, Pietra M, et al. Use of the vertebral heart score in coughing dogs with chronic degenerative mitral valve disease. *J Vet Med Sci* 2009;71:9–13.
13. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *J Am Med Assoc* 1994;271:703–707.
14. Lamb CR. Statistical briefing: Likelihood ratios. *Vet Radiol Ultrasound* 2007;48:394–395.
15. Akobeng AK. Understanding diagnostic tests 2: Likelihood ratios, pre- and post-test probabilities and their use in clinical practice. *Acta Paediatr* 2007;96:487–491.
16. Gill CJ, Sabin L, Schmid CH. Why clinicians are natural bayesians. *Br Med J* 2005;330:1080–1083.
17. Hansson K, Häggström J, Kwart C, et al. Interobserver variability of vertebral heart size measurements in dogs with normal and enlarged hearts. *Vet Radiol Ultrasound* 2005;46:122–130.
18. Buchanan JW, Bucheler J. Vertebral scale system to measure canine heart size in radiographs. *J Am Vet Med Assoc* 1995;206:194–199.
19. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307–310.
20. Matthews JN. A refinement to the analysis of serial data using summary measures. *Stat Med* 1993;12:27–37.
21. Parker CB, DeLong ER. ROC methodology within a monitoring framework. *Stat Med* 2003;22:3473–3488.
22. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44:837–845.
23. Conventional ROC methodology. *Journal of the ICRU* 2008;8:23–27.
24. Brown MD, Reeves MJ. Evidence-based emergency medicine/skills for evidence-based emergency care. Interval likelihood ratios: Another advantage for the evidence-based diagnostician. *Ann Emerg Med* 2003;42:292–297.
25. Sonis J. How to use and interpret interval likelihood ratios. *Fam Med* 1999;31:432–437.
26. Coste J, Jourdain P, Pouchot J. A gray zone assigned to inconclusive results of quantitative diagnostic tests: Application to the use of brain natriuretic peptide for diagnosis of heart failure in acute dyspneic patients. *Clin Chem* 2006;52:2229–2235.
27. Peirce JC, Cornell RG. Integrating stratum-specific likelihood ratios with the analysis of ROC curves. *Med Decis Making* 1993;13:141–151.
28. Choi BC. Slopes of a receiver operating characteristic curve and likelihood ratios for a diagnostic test. *Am J Epidemiol* 1998;148:1127–1132.
29. Simel DL, Samsa GP, Matchar DB. Likelihood ratios for continuous test results—making the clinicians' job easier or harder? *J Clin Epidemiol* 1993;46:85–93.
30. Hayden SR, Brown MD. Likelihood ratio: A powerful tool for incorporating the results of a diagnostic test into clinical decisionmaking. *Ann Emerg Med* 1999;33:575–580.
31. Fagan TJ. Letter: Nomogram for Bayes theorem. *N Engl J Med* 1975;293:257.
32. Attia JR, Nair BR, Sibbritt DW, et al. Generating pre-test probabilities: A neglected area in clinical decision making. *Med J Aust* 2004;180:449–454.
33. Halkin A, Reichman J, Schwaber M, et al. Likelihood ratios: getting diagnostic testing into perspective. *QJM* 1998;91:247–258.
34. Lamb CR. Statistical briefing: SpPinS and SnNOuts. *Vet Radiol Ultrasound* 2007;48:486–487.
35. Pewsner D, Battaglia M, Minder C, et al. Ruling a diagnosis in or out with “SpPin” and “SnNOut”: A note of caution. *Br Med J* 2004;329:209–213.
36. Kirberger RM, Lobetti RG. Radiographic aspects of *Pneumocystis carinii* pneumonia in the Miniature Dachshund. *Vet Radiol Ultrasound* 1998;39:313–317.
37. Lamb CR, Wikeley H, Boswood A, et al. Use of breed-specific ranges for the vertebral heart scale as an aid to the radiographic diagnosis of cardiac disease in dogs. *Vet Rec* 2001;148:707–711.
38. Kraetschmer S, Ludwig K, Meneses F, et al. Vertebral heart scale in the Beagle dog. *J Small Anim Pract* 2008;49:240–243.
39. Lamb CR. Statistical briefing: Accuracy and precision. *Vet Radiol Ultrasound* 2008;49:600–601.
40. Greco A, Meomartino L, Raiano V, et al. Effect of left versus right recumbency on the vertebral heart score in normal dogs. *Vet Radiol Ultrasound* 2008;49:454–455.