Repeatability and Intra- and Inter-observer Agreement of Cervical Vertebral Sagittal Diameter Ratios in Horses with Neurological Disease

K.J. Hughes, E.H. Laidlaw, S.M. Reed, J. Keen, J.B. Abbott, T. Trevail, G. Hammond, T.D.H. Parkin, and S. Love

Background: Sagittal ratio values (SRVs) of cervical vertebrae are used for ante-mortem diagnosis of cervical vertebral stenotic myelopathy, but intraobserver and interobserver variability in measurement may influence radiographic interpretation of vertebral stenosis in horses with neurological disease.

Objectives: To determine intraobserver repeatability in SRVs, intra- and interobserver agreement in SRVs and whether or not agreement was influenced by animal age.

Animals: Forty-two horses (>1 year old) with neurological disease from which laterolateral computed radiographic images of C2–C7 were obtained.

Methods: Four observers made measurements from C2 to C7 for each horse and interobserver agreement for intra- and intervertebral SRVs was determined using Bland–Altman analysis (acceptable agreement: limits of agreement [LOA] ≤ 0.05) on all horses and those ≤3 (n = 25) and >3 (n = 17) years old. Each observer also made repeated measurements for 10 horses and intraobserver repeatability and agreement were determined.

Results: Adequate intraobserver repeatability was achieved for 6 sites. Within observers, paired measurements had a median difference ≤5.7%, but a large range in differences often occurred, most frequently at intervertebral sites. For C5, C6, C7, and C3–4, LOA ≤ 0.05 were achieved by at least 1 observer. With the exception of C5 for 1 pair, LOA were >0.05 for interobserver agreement, regardless of animal age. LOA were largest at intervertebral sites.

Conclusions and Clinical Importance: Within and between observers, measurement error may limit the diagnostic accuracy of SRVs and result in discrepancies of diagnosis and treatment and warrants consideration when used clinically in horses with neurological disease.

Key words: Equine; Measurement error; Radiography.

Cervical vertebral stenotic myelopathy (CVSM) is a common cause of spinal ataxia and upper motor neuron paresis in horses and results from stenosis of the cervical vertebral canal and extradural spinal cord compression.[1] The disease occurs most commonly in young, rapidly growing horses. Certain breeds, including

Abbreviations:

CR	computed radiography
CVSM	cervical vertebral stenotic myelopathy
DICOM	digital imaging and communication in medicine
LOA	limits of agreement
RC	repeatability coefficient
SRV	sagittal ratio value

Thoroughbreds, Warmbloods, Tennessee Walking Horses and Draft breeds appear predisposed, and male horses are at increased risk for CVSM.[2–4] Two forms of CVSM have been described. Type 1 is caused by vertebral malformation and malarticulation leading to dynamic instability of the vertebral canal and is most common in young horses, whereas type 2 occurs in older horses and results from cervical osteoarthropathy leading to static compression of the spinal cord.[2,5]

Confirmation of CVSM is obtained by gross and histological postmortem examination or myelography,[1] but less invasive ante-mortem methods are required for clinical diagnosis. Accurate identification of vertebral canal stenosis is required for diagnosis of CVSM, especially when surgical intervention is considered. Although qualitative assessment of cervical radiographs for bony malformation may be suggestive of CVSM,[6–8] sensitivity and specificity are inadequate for accurate diagnosis.[6,8] To avoid the subjectivity of qualitative assessment and maximize detection of stenosis without the confounding effects of variable radiographic magnification, minimum sagittal ratio values (SRVs) of the vertebral canal from lateral radiographs commonly are

used for diagnosis of CVSM.[6,7,9] In 1 study, the sensitivity and specificity of intravertebral SRVs for diagnosis of CVSM in young horses (1–4 years of age) were ≥89%, but the site or sites of cord compression could not be identified specifically.[6] More recently, lower sensitivity (47%) and specificity (78%) of intravertebral SRVs for diagnosis of CVSM in older horses (≥4 years old) was found,[5] suggesting an influence of animal age on method accuracy. In another study,[9] intervertebral SRVs were found to be accurate for the diagnosis of CVSM and identification of the compression site, but only 8 horses with CVSM were examined.

In addition to the importance of the sensitivity and specificity of SRVs for diagnosis of CVSM, variability of measurements, both within and between observers, may have implications for the accuracy of the method. Recently, in a study of observer agreement of SRVs in horses, variability in SRVs of up to 10%, both within and between observers, was found.[10] In that study, measurements were made at C3–4 and C6–7 as representative sites for compression in young and older horses, respectively.[10] However, spinal cord compression in horses with CVSM also may occur at C2–3, C4–5, C5–6 or some combination of these sites,[6,8] intra- and intervertebral SRVs from C2 to C7 inclusive have been described,[5,6,9] and observer agreement of SRV measurements at these sites is unknown. The objectives of this study were to determine intraobserver repeatability in intra- and intervertebral SRVs, intra- and interobserver agreement in intra- and intervertebral SRVs for C2–C7 in horses, and whether agreement was influenced by animal age.

Materials and Methods

Study Population and Case Selection

Medical records and archived computed radiographs of all horses ≥1 year old with neurological disease that had undergone computed radiography (CR) of the cervical vertebrae between June 2007 and June 2009 at 1 of 3 institutions (University of Glasgow, Rood and Riddle Equine Hospital and Virginia-Maryland Regional College of Veterinary Medicine) were reviewed. The criteria for inclusion were availability of signalment data (age, breed, sex) and a radiographic series of the cervical vertebrae, comprised of plain, laterolateral images of C2–C7 inclusive. Radiographic series were reviewed by a board-certified specialist in equine medicine (KJH) and a veterinary student (EHL) and series in which ≥1 of C2–C7 was not imaged, radiographic quality was considered inadequate, obliquity was present or some combination of these were excluded from the study. For each recruited animal, outcome data and diagnosis, when known, were recorded. For this study, CVSM was diagnosed based on neurologic examination findings and ≥1 of the following: SRVs, myelographic results and gross and histological postmortem examination findings of spinal cord compression.

The study was approved by the Ethics and Welfare committee, School of Veterinary Medicine, University of Glasgow.

Radiography

Archived computed radiographs from each of the 3 institutions were used for the study. All data were stored in digital imaging and communication in medicine (DICOM) standard format and images were reviewed by the observers on computer screens.

Observers

Radiographic measurements were made by board-certified specialists in equine medicine (JK and JA) and diagnostic imaging (GH) and a diagnostic imaging resident (TT). All observers had experience in the assessment of equine cervical vertebral radiographs, but observers 1 and 2 had more experience in the acquisition and interpretation of SRVs than did observers 3 and 4.

Radiographic Measurements

All observers were sent the complete series of laterolateral cervical radiographs for each horse. When a specific vertebra appeared >1 radiograph in a series, the observers were instructed as to which radiograph was to be used for the intravertebral, intervertebral or both measurements. All measurements were made using electronic calipers included in the software provided to observers for viewing the DICOM files. For each horse, the following measurements were made for calculation of the intravertebral SRVs[6] for C2–C7 inclusive and intervertebral SRVs[9] for C2–3 to C6–7 inclusive (Fig 1):

1 minimum sagittal diameter (a): measured at any point along the vertebral canal at the minimum diameter
2 maximum height of the cranial vertebral epiphysis (b)
3 intervertebral minimum sagittal diameter (c): the shortest distance from the caudal dorsal lamina of the more cranial vertebra to the dorsal aspect of the cranial epiphysis of the more caudal vertebra (c1) or from the dorsal aspect of the caudal epiphysis of the more cranial vertebra to the cranial dorsal lamina of the more caudal vertebra (c2). The smallest value between c1 and c2 was recorded.
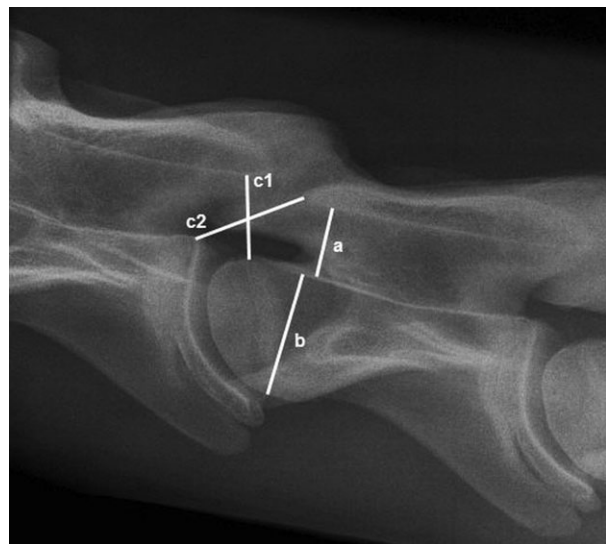
Fig 1. A laterolateral radiographic image of cervical vertebrae 3 and 4 demonstrating the intravertebral minimum sagittal diameter **(a)**, maximal height of the cranial vertebral physis **(b)**, and intervertebral minimum sagittal diameter measurements **(c1, c2)** made to calculate intravertebral **(a/b)** and intervertebral **(c/b)** sagittal ratio values (SRVs). For each intervertebral SRV calculation, the smaller value of **c1** and **c2** was used.

All measurements were recorded by each observer in an electronic data sheet[a] which was used by one author (EHL) to calculate the intra- and intervertebral SRVs. This part of the study was designed to determine interobserver agreement in SRV measurements.

Two weeks later, 10 radiographic series were selected randomly from the study population using a random number generator[a] and were sent to the observers for repeated measurements to determine intraobserver agreement and repeatability of both intra- and intervertebral SRV measurements.

Blinding

The radiographic images were retained in DICOM format to preserve image resolution, maximize accuracy of measurements and enable use of the electronic calipers incorporated into the software for collection of measurements. Because the DICOM software used did not allow application of anonymity to the images, observers were not blinded to horse identity during intraobserver variability assessment. The intervening period of 2 weeks for the intraobserver study was selected to minimize observer bias. Observers were blinded to clinical and outcome data of the horses. Each observer was unaware of the results of the other observers.

Statistical Analysis

All analyses were performed using commercially available software.[a] Intraobserver repeatability of intra- and intervertebral SRVs at each site was determined by comparing the absolute differences between repeated measurements to the standard deviation (SD) of the differences of the paired measurements.[11] Repeatability was assumed when ≥90% of the absolute differences in the paired measurements were <1.96 SD of the differences (repeatability coefficient [RC]). Intraobserver repeatability also was examined by determining the median and range difference (%) in SRV for each site for each observer. Good repeatability was assumed with a difference in SRV of <5%. Bland–Altman plots[11] were produced to examine intra- and interobserver agreement for both intra- and intervertebral SRVs at each site. Interobserver agreement was assessed for the whole study population, horses ≤3 and >3 years of age. Acceptable agreement was considered to be present when the limits of agreement (LOA) on the Bland–Altman plots were ≤0.05 (5%). To determine whether interobserver variation in SRV measurement at each site was associated with age (≤3 years, >3 years), 2-tail, 2-sample t-tests were performed, using the absolute differences in paired measurements for each observer pair and significance was set at $P < .05$. Using neurologic postmortem examination results as the gold standard test for CVSM, the sensitivity and specificity of SRVs at both intravertebral sites and intervertebral sites were calculated for each observer. For determination of sensitivity and specificity, published cut-off values[1,6,9] were used: a positive test result was recorded if ≥1 of the SRVs were less than the cut-off values and a negative test result was recorded if all of the SRVs were more than the cut-off values.

Results

Study Population

A total of 73 CR series of cervical vertebrae from horses with neurological disease were reviewed. After review, 31 series were excluded, resulting in the inclusion of 42 horses for the study.

The study population consisted of 33 male and 9 female horses. Median age at time of radiography was 3 years (range, 1–18; interquartile range, 3–7). Twenty-five horses were ≤3 years old and 17 horses were >3 years old. There were 21 Thoroughbreds, 6 Quarter Horses, 5 Warmbloods, 3 American Saddlebreds, 3 Thoroughbred crosses, 2 Tennessee Walking Horses, 1 Kentucky Mountains Saddle Horse and 1 Irish Sport Horse. Twenty-nine horses were diagnosed with CVSM (20 by postmortem examination, 3 by myelography, and 6 by SRVs). For the 29 horses with CVSM, there were 23 male and 6 female horses. Seven horses in the study had equine protozoal myeloencephalitis, diagnosed by immunological testing (*Sarcocystis neurona*, either by western blot test, surface antigen ELISA or indirect fluorescent antibody test) of cerebrospinal fluid samples and serum (2 horses), postmortem examination (2 horses) or immunological testing and postmortem examination (3 horses). One of the 7 horses with equine protozoal myeloencephalitis also had gross and histological changes of CVSM. One horse had lower motor neuron disease, diagnosed by postmortem examination. A diagnosis was not available for 5 horses. A postmortem examination was performed on 1 of the 5 horses for which a diagnosis was not available and multifocal hemorrhagic lesions were present in the cervical spinal cord but no additional testing was performed.

Repeatability

Observers 1 and 2 had acceptable repeatability (RC, ≥90%) at all intravertebral and intervertebral sites examined (Table 1). A RC of <90% was obtained at 3 sites and 1 site for observers 3 and 4, respectively. Overall, C2–3, C4, C4–5, C5–6, C7, and C6–7 had

Table 1. Intraobserver repeatability coefficients (%) for cervical intravertebral and intervertebral sites.

Observer	Vertebral Site									
	C3	C2–3	C4	C3–4	C5	C4–5	C6	C5–6	C7	C6–7
1	90	90	90	90	90	90	90	90	100	90
2	90	90	90	100	100	100	100	100	90	90
3	80[a]	90	90	20[a]	70[a]	90	90	90	90	100
4	90	90	90	100	100	100	80[a]	100	90	100
Overall average	87.5[a]	90	90	77.5[a]	87.5[a]	95	90	95	92.5	95

[a]Repeatability coefficient less than the criterion of acceptability for the study (90%).

acceptable repeatability across the 4 observers. When the requirement of a RC of 100% was applied, no site had acceptable repeatability overall, whereas repeatability was achieved at 5 sites for observers 2 and 4, and at 1 site for observers 1 and 3. The median difference in paired SRVs by site ranged from 1.2 to 4.7%, 1.2 to 4.9%, 1.1 to 3.8%, and 1.2 to 5.8% for observers 1, 2, 3, and 4, respectively (Table 2). Despite the limited spread in median differences, the largest absolute ranges were 0.6–26.5%, 0.6–27.7%, 1.9–18.0%, and 0.7–26.6% for observers 1, 2, 3, and 4, respectively. Overall, and for each observer, the median and range of differences in SRVs were most often higher for intervertebral sites than for intravertebral sites.

Intraobserver Agreement

Mean bias and LOA data for the intravertebral and intervertebral sites are presented in Table 3. For C5, C6, C7, and C3–4, at least 1 observer achieved the criterion for agreement, whereas for all other sites, no observer achieved acceptable agreement. Overall, across all observers, agreement was most often achieved for C5.

Interobserver Agreement

Mean bias and LOA data for the intra- and intervertebral sites are provided in Tables 4, 5, respectively. For both intra- and intervertebral sites, agreement was not achieved for all horses and for horses aged <3 years, for any observer pair. For all repeated measurements for horses aged >3 years, agreement was achieved on 1 occasion (C5, observer pair 1&3; Table 4). For intravertebral sites, some LOA (eg, for C6) were close to the criterion for agreement. For intervertebral sites, the LOA ranges often were large, representing poor interobserver agreement (Table 5). For each observer pair, there was no significant difference in the absolute differences in paired measurements at each site for horses aged ≤3 and >3 years. In general, for each observer pair, measurement differences did not vary over the range of mean measurements for more cranial sites (ie, C3–C5; Fig 2), whereas for more caudal sites (C5–C7) there often was an increase in the scatter of the differences as the mean increased (Fig 3).

Sensitivity and Specificity of SRVs

A neurological postmortem examination was performed on 27 horses in the study population. Twenty horses had CVSM and 7 horses did not have CVSM (EPM [n = 4], LMND [n = 1], other [n = 2]). Sensitivity and specificity results of intravertebral and intervertebral SRVs for each observer are provided in Table 6. Using intravertebral ratios of <50% for C3–C6 and <52% for C7 as diagnostic for CVSM,[1] median sensitivity and specificity were 69% and 61%, respectively. When ratios of <52% for C3–C6 and <56% for C7 were used for diagnosis of CVSM,[6] median sensitivity and specificity were 84% and 32%, respectively. For

Table 2. Median (range) intraobserver differences in sagittal ratio value for intravertebral and intervertebral sites.

Observer	C3	C2–3	C4	C3–4	C5	C4–5	C6	C5–6	C7	C6–7
					Vertebral Site					
1	2.9% (0.0–17.4)	4% (0.0–11.1)	2.2% (0.0–6.0%)	4.7% (0.0–12.1)	1.2% (0.0–24.2)	2.2% (0.0–17.9)	1.7% (0.0–6.4%)	3.1% (0.0–15.2)	3.9% (0.0–7.7)	3.9% (0.6–26.5)
2	2.2% (0.3–23.4)	3.5% (0.6–27.7)	1.5% (1.0–8.9)	1.8% (0.0–4.4)	2.2% (0.0–4.8)	2.2% (0.8–6.5)	1.2% (0.0–5.5)	4.9% (0.4–16.4)	1.8% (0.0–4.8)	3.9% (1.3–12.0)
3	1.8% (0.0–11.9)	3.6% (1.9–18.0)	3.3% (1.1–11.9)	3.8% (0.0–6.8)	1.1% (0.3–4.5)	1.2% (0.4–9.4)	1.1% (0.0–4.1)	2.5% (0.0–15.2)	1.8% (0.2–14.3)	2.6% (0.7–5.6)
4	2.1% (0.0–10.7)	5.8% (0.0–22.9)	2.8% (0.0–6.8)	5.7% (0.0–10.1)	1.2% (0.0–3.5)	2.3% (0.7–8.5)	2.6% (0.5–9.8)	4.1% (2.2–11.6)	3.3% (0.0–10.7)	4.2% (0.7–26.6)

Table 3. Intraobserver agreement for intravertebral and intervertebral SRVs.

Vertebral Site

Observer	C3 Mean bias	C3 LOA Lower	C3 LOA Upper	C4 Mean bias	C4 LOA Lower	C4 LOA Upper	C5 Mean bias	C5 LOA Lower	C5 LOA Upper	C6 Mean bias	C6 LOA Lower	C6 LOA Upper	C7 Mean bias	C7 LOA Lower	C7 LOA Upper
1	0.01	−0.11	0.14	−0.01	−0.07	0.04	0.04	−0.12	0.19	−0.01	−0.06	0.05	−0.01	−0.12	0.09
2	−0.03	−0.19	0.12	0.01	−0.05	0.07	0.00	−0.05	0.05[a]	−0.01	−0.07	0.05	−0.01	−0.05	0.04[a]
3	0.02	−0.09	0.13	0.01	−0.09	0.11	0.02	−0.01	0.05[a]	0.01	−0.02	0.04[a]	0.02	−0.09	0.12
4	0.00	−0.08	0.09	0.02	−0.05	0.08	0.01	−0.02	0.04[a]	0.03	−0.05	0.10	0.01	−0.09	0.11

Observer	C2–3 Mean bias	C2–3 LOA Lower	C2–3 LOA Upper	C3–4 Mean bias	C3–4 LOA Lower	C3–4 LOA Upper	C4–5 Mean bias	C4–5 LOA Lower	C4–5 LOA Upper	C5–6 Mean bias	C5–6 LOA Lower	C5–6 LOA Upper	C6–7 Mean bias	C6–7 LOA Lower	C6–7 LOA Upper
1	0.00	−0.11	0.10	−0.03	−0.15	0.09	0.00	−0.17	0.17	0.03	−0.10	0.15	−0.02	−0.22	0.17
2	0.01	−0.19	0.20	0.01	−0.03	0.05[a]	−0.01	−0.10	0.07	0.01	−0.12	0.15	0.02	−0.09	0.14
3	0.06	−0.06	0.19	0.04	−0.00	0.07	0.01	−0.07	0.10	0.02	−0.09	0.14	0.02	−0.04	0.08
4	−0.03	−0.23	0.18	−0.02	−0.14	0.10	−0.01	−0.11	0.08	0.01	−0.12	0.13	−0.03	−0.24	0.17

SRVs, sagittal ratio values; LOA, limits of agreement.
[a]Measurement that meets the defined criterion for agreement (LOA ≤ 0.05).

intervertebral sites, a cut-off value of <48.5%[9] resulted in median sensitivity of 20% and median specificity of 100%.

Discussion

This study was undertaken to assess the repeatability and agreement between paired SRV measurements, both within and between observers, for all cervical sites that may be affected in CVSM. It is important to establish the agreement (and conversely variability) in SRV measurements to determine whether these measurements are suitable and reliable for clinical use. The results show that intraobserver repeatability for measurement of intra- and intervertebral SRVs from equine cervical radiographs is moderate. However, some intraobserver variability in measurements occurred and often resulted in failure to achieve the definition of agreement for paired SRV measurements, because repeatability influences the amount of achievable agreement.[11] This intraobserver variability also influenced the interobserver agreement achieved, because limitations in agreement are amplified when comparing 2 observers or methods with less than perfect repeatability.[11] As such, considerable interobserver variation was present for SRV measurements at intra- and intervertebral sites, and agreement was not achieved for any observer pair, with the exception of C5 for observers 1 and 3 when horses aged >3 years were considered (Table 4). These findings suggest the limitations in measurement repeatability and agreement will influence the clinical reliability of SRVs for diagnosis or exclusion of CVSM, particularly for different observers because there is insufficient agreement to permit interchanging of measurements. Similarly, in an earlier study, more variation in SRV measurements was found between 2 observers than within a single observer and agreement was not achieved for any interobserver comparison.[10] However, only C4, C3–4, C7, and C6–7 sites were measured in that study, and the current study is the first to determine agreement at all sites that may be involved in CVSM.[1,4,6]

Overall, when all observers were considered, acceptable repeatability was achieved for SRV measurements at 7 of 10 cervical sites examined (C2–3, C4, C4–5, C6, C5–6, C7, and C6–7; Table 1). The RC of ≥90% in the current study is less than that recommended for determination of repeatability (95%)[11] and was chosen because of the limited number of paired observations per site for each observer (n = 10). When a more rigorous criterion (RC = 100%) was applied, no site had acceptable repeatability overall, although 2 observers had repeatability at 1 site and the 2 more experienced observers had repeatability at 5 sites. In an earlier study, repeatability (RC of ≥95%) was achieved at C4, C3–4 and C7 but not at C6–7, when repeated observations were made from 75 radiographic sets by a single observer.[10] In that study, the RC for C6–7 was 92%,[10] whereas in the current study, the overall coefficient was 95%, and 2 observers had a coefficient of 100%. The authors of the earlier study speculated that the

Table 4. Interobserver agreement for intravertebral SRVs for all 42 horses (All), horses ≤3 years of age (n = 25) and horses >3 years of age (n = 17).

		Site														
		C3			C4			C5			C6			C7		
			LOA			LOA			LOA			LOA			LOA	
Observers		Mean bias	Lower	Upper	Mean bias	Lower	Upper	Mean bias	Lower	Upper	Mean bias	Lower	Upper	Mean bias	Lower	Upper
1&2	All	0.04	−0.08	0.15	0.05	−0.05	0.15	0.04	−0.06	0.13	0.04	−0.04	0.11	0.05	−0.05	0.15
	≤3 years	0.05	−0.09	0.18	0.05	−0.05	0.16	0.04	−0.07	0.15	0.03	−0.02	0.08	0.07	−0.03	0.17
	>3 years	0.03	−0.03	0.08	0.04	−0.06	0.14	0.03	−0.03	0.10	0.04	−0.05	0.14	0.03	−0.06	0.12
1&3	All	0.01	−0.06	0.09	0.02	−0.07	0.11	0.01	−0.06	0.08	0.01	−0.06	0.08	0.01	−0.08	0.10
	≤3 years	0.02	−0.07	0.11	0.02	−0.06	0.10	0.01	−0.07	0.10	0.01	−0.07	0.09	0.02	−0.04	0.09
	>3 years	0.01	−0.04	0.06	0.01	−0.10	0.13	0.00	−0.05	0.05[a]	0.00	−0.06	0.06	−0.01	−0.13	0.10
1&4	All	0.04	−0.03	0.11	0.05	−0.03	0.14	0.04	−0.04	0.13	0.03	−0.04	0.09	0.03	−0.05	0.12
	≤3 years	0.04	−0.03	0.11	0.06	−0.02	0.14	0.05	−0.05	0.15	0.03	−0.03	0.09	0.04	−0.04	0.11
	>3 years	0.03	−0.05	0.11	0.04	−0.04	0.13	0.03	−0.02	0.09	0.03	−0.05	0.11	0.03	−0.08	0.13
2&3	All	0.03	−0.08	0.13	0.03	−0.09	0.15	0.03	−0.05	0.11	0.03	−0.07	0.12	0.04	−0.09	0.18
	≤3 years	0.03	−0.09	0.14	0.04	−0.07	0.14	0.02	−0.07	0.12	0.02	−0.07	0.10	0.05	−0.07	0.16
	>3 years	0.02	−0.07	0.11	0.03	−0.11	0.16	0.03	−0.02	0.09	0.04	−0.07	0.15	0.04	−0.12	0.20
2&4	All	0.00	−0.10	0.11	0.00	−0.09	0.08	−0.01	−0.08	0.07	0.01	−0.09	0.11	0.02	−0.09	0.13
	≤3 years	0.00	−0.12	0.13	0.00	−0.10	0.09	−0.01	−0.10	0.07	0.00	−0.07	0.08	0.03	−0.08	0.15
	>3 years	0.00	−0.07	0.07	0.00	−0.07	0.06	0.00	−0.06	0.06	0.01	−0.13	0.15	0.00	−0.09	0.10
3&4	All	−0.02	−0.11	0.07	−0.04	−0.12	0.05	−0.03	−0.12	0.06	−0.02	−0.10	0.06	−0.02	−0.13	0.09
	≤3 years	−0.02	−0.11	0.06	−0.04	−0.12	0.04	−0.03	−0.14	0.07	−0.01	−0.08	0.06	−0.01	−0.09	0.06
	>3 years	−0.02	−0.13	0.08	−0.03	−0.12	0.06	−0.03	−0.10	0.04	−0.03	−0.11	0.05	−0.03	−0.18	0.11

SRVs, sagittal ratio values; LOA, limits of agreement.
[a]Measurement that meets the defined criterion for agreement (LOA ≤ 0.05).

Table 5. Interobserver agreement for intervertebral SRVs for all 42 horses (All), horses ≤3 years of age (n = 25) and horses >3 years of age (n = 17).

		Site														
		C2–3			C3–4			C4–5			C5–6			C6–7		
			LOA			LOA			LOA			LOA			LOA	
Observers		Mean bias	Lower	Upper	Mean bias	Lower	Upper	Mean bias	Lower	Upper	Mean bias	Lower	Upper	Mean bias	Lower	Upper
1&2	All	0.06	−0.17	0.29	0.01	−0.12	0.14	0.03	−0.14	0.21	0.05	−0.11	0.21	0.06	−0.19	0.30
	≤3 years	0.08	−0.18	0.35	0.01	−0.13	0.15	0.04	−0.18	0.26	0.02	−0.09	0.14	0.08	−0.10	0.26
	>3 years	0.03	−0.14	0.19	0.00	−0.11	0.12	0.03	−0.05	0.11	0.09	−0.10	0.28	0.07	−0.07	0.21
1&3	All	−0.07	−0.25	0.11	−0.09	−0.23	0.06	−0.08	−0.19	0.03	−0.05	−0.18	0.13	0.01	−0.14	0.16
	≤3 years	−0.06	−0.24	0.12	−0.09	−0.23	0.05	−0.07	−0.19	0.05	−0.07	−0.19	0.05	0.01	−0.14	0.16
	>3 years	−0.08	−0.27	0.11	−0.08	−0.24	0.07	−0.08	−0.18	0.01	−0.03	−0.17	0.11	0.01	−0.15	0.16
1&4	All	0.08	−0.08	0.23	0.04	−0.10	0.18	0.04	−0.10	0.17	0.04	−0.10	0.18	0.03	−0.05	0.12
	≤3 years	0.08	−0.08	0.25	0.04	−0.09	0.17	0.04	−0.11	0.20	0.02	−0.08	0.12	0.07	−0.07	0.21
	>3 years	0.07	−0.07	0.21	0.04	−0.12	0.20	0.02	−0.08	0.13	0.07	−0.10	0.24	0.07	−0.17	0.32
2&3	All	0.13	−0.15	0.42	0.09	−0.05	0.24	0.12	−0.02	0.23	0.10	−0.06	0.26	0.07	−0.09	0.23
	≤3 years	0.14	−0.19	0.48	0.10	−0.07	0.26	0.11	−0.04	0.26	0.09	−0.08	0.25	0.08	−0.08	0.23
	>3 years	0.11	−0.08	0.30	0.09	−0.02	0.20	0.12	−0.01	0.26	0.12	−0.03	0.26	0.07	−0.10	0.23
2&4	All	−0.02	−0.30	0.26	−0.03	−0.16	0.10	0.00	−0.15	0.15	0.01	−0.17	0.18	0.00	−0.22	0.21
	≤3 years	0.00	−0.30	0.30	−0.03	−0.16	0.10	−0.01	−0.17	0.16	0.00	−0.14	0.13	0.00	−0.22	0.23
	>3 years	−0.05	−0.29	0.20	−0.03	−0.18	0.11	0.01	−0.12	0.14	0.02	−0.20	0.25	−0.01	−0.21	0.20
3&4	All	−0.14	−0.35	0.06	−0.13	−0.31	0.06	−0.11	−0.22	−0.01	−0.09	−0.23	0.05	−0.02	−0.13	0.09
	≤3 years	−0.14	−0.35	0.06	−0.13	−0.29	0.04	−0.12	−0.21	−0.03	−0.09	−0.25	0.07	−0.06	−0.22	0.10
	>3 years	−0.15	−0.37	0.07	−0.12	−0.33	0.09	−0.11	−0.24	0.01	−0.09	−0.21	0.02	−0.06	−0.25	0.12

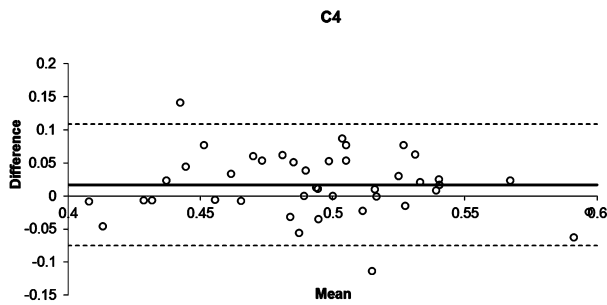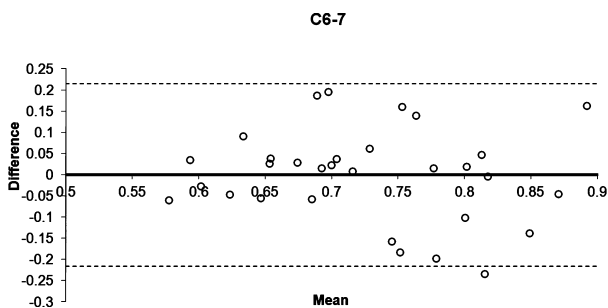SRVs, sagittal ratio values; LOA, limits of agreement.

Fig 2. Bland–Altman plot of the differences of intravertebral sagittal ratio value (SRV) measurements for C4 between observers 1 and 3. The x-axis is the mean of the paired measurements for each horse and the y-axis is the difference between the paired measurements for each horse. The solid line depicts the mean bias and the dashed lines depict the upper and lower limits of agreement (LOA). The scatter of the differences does not change as the mean SRV measurement increases, and the calculated LOA are expected to have good accuracy.

Fig 3. Bland–Altman plot of the differences of intervertebral sagittal ratio value (SRV) measurements for C6–7 between observers 2 and 4. The x- and y-axes are the same as Figure 2. The scatter of the differences increases as the mean SRV measurement increases, and the limits of agreement (dashed lines) will be wider and narrower than necessary for small and large SRV measurements, respectively.

lower RC for C6–7, compared to the other sites examined, may have been because of radiograph underexposure and inability to accurately define anatomical landmarks because of superimposition of the thoracic limbs.[10] In comparison to that study,[10] the capacity for postacquisition imaging processing afforded by the use of DICOM images in the current study may have contributed to the higher overall repeatability at C6–7. However, the limited number of paired observations for determination of intraobserver repeatability (and agreement) is a limitation of the current study, and our findings need to be interpreted accordingly, because repeatability may have been over-estimated. When repeatability was assessed by examination of the absolute differences in paired observations, the median difference in SRV measurements for each site for each observer was ≤5.8%, and 60% were <3.0% (Table 2). Although these findings suggest good repeatability, the range in paired differences (0–27.7%) indicates that, on occasion, repeatability was poor at individual sites.

Repeatability is important because it will limit the amount of intraobserver agreement achievable. Accordingly, for individual observers, sites with the smallest median and range differences for paired observations were more likely to achieve or approach the criterion for agreement (LOA ≤ 0.05).

In the current study, variability within and between observers for measurements of vertebral body diameter ("b") and vertebral canal diameters ("a," "c1," and "c2") from the same radiographic image may occur for several reasons. Differences in caliper placement are inevitable because of the requirement for visual determination of the narrowest and widest points of the vertebral canal and cranial aspect of the vertebral body, respectively. Interpretation of the appropriate location for caliper placement may have differed among observers. For example, interpretation of the interface of the ventral aspect of the vertebral canal and dorsal aspect of the vertebral body may differ, resulting in differences in caliper placement for "a" and "b" measurements. Furthermore, software limitations in the precise caliper positioning in relation to the anatomical margins of the vertebrae may have contributed to intra- and interobserver variability. However, studies comparing digital and conventional radiography in human dentistry found no differences in linear measurements,[12,13] suggesting caliper placement is likely to be as accurate as manual measurement on screen film. In our study, measurement "a" was determined at any point along the vertebral canal where the diameter was considered minimum,[9,14] whereas in the study of Scrivani et al[10] the intravertebral canal measurement was made at the cranial aspect.[6] The method used in our study was chosen because it is used clinically,[9,14] but it may have resulted in more intra- and interobserver variability of paired "a" measurements and greater variability in SRV values, compared to the earlier study.[10] Similarly, differences exist between the 2 studies in collection of intervertebral measurements. In the earlier study,[10] the minimum distance between the caudal aspect of the lamina of the more cranial vertebra and the cranial aspect of the epiphysis of the more caudal vertebra was used, whereas we used the smaller of "c1" and "c2" measurements (Fig 1), as described previously.[9] The selection of either "c1" or "c2" may have contributed to the poorer interobserver agreement in intervertebral SRVs in our study compared to the previous study.[10] Similarly, the use of "c1" or "c2" may have influenced the poorer agreement we found for intervertebral SRVs compared to intravertebral SRVs, both within (Table 3) and among observers (Tables 4, 5). A final consideration for limitations in repeatability and agreement in SRVs is the potential for amplification of measurement variability because of the requirement for 2 measurements to determine the ratio. Although the agreement in absolute measurements may have been higher than for derived SRVs in the current and previous study,[10] this is not of clinical relevance because of the reliance on ratio values to account for magnification of absolute measurements and to maximize the

Table 6. Sensitivity and specificity of SRVs at intravertebral and intervertebral sites in horses with (n = 20) and without (n = 7) CVSM confirmed by neurological postmortem examination.

| | Intravertebral Sites[1,6] | | | | Intervertebral Sites[9] | |
| | Method 1[1] | | Method 2[6] | | | |
Observer	Sn (%)[a]	Sp (%)[b]	Sn (%)[a]	Sp (%)[b]	Sn (%)[a]	Sp (%)[b]
1	85	29	90	14	20	100
2	55	86	75	43	15	100
3	70	29	95	14	40	100
4	65	100	75	57	5	100
Median	69	61	84	32	20	100

CVSM, cervical vertebral stenotic myelopathy; Sn, sensitivity; Sp, specificity; SRV, sagittal ratio values.

[a]A positive test result was recorded if one or more of the SRVs were less than published cut-off values.

[b]A negative test result was recorded if all of the SRV were greater than published cut-off values.

diagnostic accuracy of the procedure.[6] The poorer intraobserver and interobserver agreement of intervertebral SRVs, compared to intravertebral SRVs, found in the current and previous study,[10] has implications for the use of this measurement for the diagnosis of CVSM. In a study of 26 ataxic horses, of which 8 were diagnosed with CVSM by histological examination, intervertebral SRVs were found to be accurate for diagnosis of CVSM and specific for identifying the site or sites of spinal cord compression.[9] However, the authors concluded that additional assessment using a larger data set is necessary to determine the sensitivity, specificity and predictive values of intervertebral SRVs for diagnosis of CVSM and site of compression,[9] and our results suggest that poor observer agreement at these sites may limit the diagnostic performance of the method.

In the current study, CR images in DICOM format and electronic calipers were used by each operator to obtain measurements. In contrast, conventional radiography (screen film) and plastic callipers were used in the study by Scrivani et al.[10] Spatial resolution of conventional radiography is superior to that of CR, because of inherent limitations in matrix and pixel size in CR systems.[15,16] The higher spatial resolution of conventional radiography may account for the reported superior observer ability to detect microcalcifications in human mammography images using this modality, in comparison to CR.[17] Higher spatial resolution of screen film images used for determination of SRVs by Scrivani et al[10] may have contributed to differences in intra- and interobserver agreement between their study and our study. However, despite lower spatial resolution of CR compared to conventional radiography, contrast resolution of CR is higher because of edge enhancement filters incorporated in the image reconstruction algorithm, ability to manipulate the image and a wider dynamic range for image processing.[15,18,19] Overall, CR often has superior image quality as compared to conventional radiography and is associated with equal or higher diagnostic performance for soft tissue and osseous lesions,[15,19–21] and it is unlikely the use of CR influenced accuracy of vertebral measurements and SRVs in our study.

Although determination of the diagnostic accuracy of SRVs was not a focus of this study, performance of the SRV method was examined briefly, using those horses that were confirmed by postmortem examination to have CVSM (n = 20) or non-CVSM neurological disease (n = 7). Using this subset of the overall study population, sensitivity and specificity of intra- and intervertebral SRVs were calculated for each observer. Overall, the sensitivity of intravertebral SRV was moderate to good, particularly when the criteria described by Moore et al[6] were used. Conversely, specificity of intravertebral SRVs was variable, and was poor for all observers using the cut-off values of Moore et al.[6] Previous authors have reported high accuracy of intravertebral SRVs for diagnosis of CVSM, with a sensitivity and specificity of ≥87%.[6,22] The lower sensitivity and specificity of intravertebral SRVs in the current study may reflect the small number of horses available for analysis and variable observer experience. In addition, most horses were ≤3 years of age in the previous study,[6] whereas there was more variability in horse age in the current study and there may be an influence of higher animal age on accuracy of SRVs, as suggested previously.[5] In our study, when a cut-off value of 48.5% was used, intervertebral SRVs were highly specific for all observers, but the method had poor sensitivity. In another study,[9] intervertebral SRVs were found to be highly accurate for the diagnosis of CVSM. Differences in CVSM lesion type[22] and observer experience are possible reasons for differences in the sensitivity of intervertebral SRVs between the 2 studies, but additional validation of the method using larger numbers of horses is required to better establish the diagnostic accuracy in horses presented with spinal ataxia.

The radiographic series represented a convenience sample, and both horses with and without CVSM were included to provide sufficient spread of SRV measurements at each vertebral site. A range in measurements is necessary to examine whether there is a consistent relationship between the difference and the mean of paired measurements or an increasing or decreasing trend.[11] This has important implications for the LOA and interpretation of agreement. For some vertebral

sites, as the mean SRV measurement increased, the scatter of the differences for observer pairs increased. In general, this trend occurred for measurements made at sites caudal to C5 and was most evident for intervertebral sites. As a consequence of this trend, the LOA were likely wider than necessary for smaller SRVs, meaning the calculated agreement was underestimated for SRVs in the range associated with CVSM and over-estimated for ranges not consistent with CVSM. The clinical implications of this finding could not be determined in this study, but it is possible that for some vertebral sites, the likelihood for a false negative diagnosis of CVSM is smaller and a false positive diagnosis is higher than suggested by our data. For other sites (eg, C3–5 inclusive), the SRV measurement differences did not, in general, vary over the range of mean measurements, suggesting the calculated LOAs provide a more accurate estimate of interobserver agreement.

It was our hypothesis that agreement would be higher for SRV measurements obtained from young horses (aged ≤3 years) because proliferative osseous changes of the dorsal articulations of cervical vertebrae, present commonly in horses aged ≥4 years,[1,5] may have resulted in more variability in intra- and intervertebral measurements and derived SRVs for horses aged >3 years. However, there was no effect of horse age on interobserver variation in SRV measurements. Despite this result, horse age still may influence the accuracy of the SRV method for detection of CVSM, based on findings from earlier studies. In a study of 100 young horses (93 aged ≤3 years) with CVSM diagnosed by myelography with or without histological confirmation and 100 age-matched controls, the sensitivity and specificity of intravertebral SRVs (C4–C7 inclusive) were ≥89%,[6] indicating the diagnostic value of the method in this age group. In a later study of horses aged ≥4 years with CVSM diagnosed by myelography, necropsy, or both and contemporaneous control horses, intravertebral SRVs (C3–C7 inclusive) had sensitivity and specificity values of 47% and 78%, respectively.[5] This latter study suggests that the SRV method only will detect approximately 50% of older horses with CVSM, which may reflect the finding that osteoarthropathy of the dorsal articulations is the most common cause of spinal cord compression in this age group (type 2 CVSM),[1,2,5] compared to sagittal vertebral canal stenosis associated with malformation in younger horses (type 1 CVSM).[1,2,6]

The results of the current study may have been influenced by observer experience. Measurements were made by 2 board-certified equine medicine specialists, 1 board-certified radiologist and a diagnostic imaging resident, and although all were familiar with the acquisition of SRVs, observers 1 and 2 were more experienced. Observer experience may influence the magnitude of achievable repeatability and intra- and interobserver agreement. A learning curve in radiographic measurement acquisition is inevitable and increased observer experience would be expected to result in higher reliability in measurements of distance, as has been reported for various applications of skeletal[23,24] and soft tissue radiography[25,26] in human medicine. In our study, the 2 observers with the most experience in cervical radiographic interpretation had higher agreement in SRV measurement than did the less experienced observers, when a RC of ≥90% was applied. In a previous study of SRVs in horses obtained by 2 observers, good intraobserver repeatability was obtained by an experienced board-certified radiologist, but repeatability was not determined for the second, inexperienced observer.[10] In that study, intraobserver agreement was superior to interobserver agreement,[10] supporting the effect of observer experience on measurement reliability.

In conclusion, intraobserver repeatability of SRV was good overall for 6 of 10 cervical vertebral sites examined. However, the often large range of paired measurement differences for each observer and only moderate intraobserver agreement indicates that within observers, important measurement error may occur. The poor agreement among observers, regardless of animal age, indicates that clinical interpretation of cervical radiographs of horses with neurological disease may differ and may be influenced by observer experience. Measurement error and variability in SRV results, both within and among observers, may limit the diagnostic accuracy of the SRV method and result in discrepancies of diagnosis and inappropriate management. Consideration of the limitations of the repeatability and agreement of SRVs is required when the results of the method are incorporated into clinical decision making when examining horses with neurological disease.

References

1. Nout YS, Reed SM. Cervical vertebral stenotic myelopathy. Equine Vet Educ 2003;15:212–223.

2. Oswald J, Love S, Parkin TD, et al. Prevalence of cervical vertebral stenotic myelopathy in a population of thoroughbred horses. Vet Rec 2010;166:82–83.

3. Levine JM, Ngheim PP, Levine GJ, et al. Associations of sex, breed, and age with cervical vertebral compressive myelopa-

thy in horses: 811 cases (1974–2007). J Am Vet Med Assoc 2008;233:1453–1458.

4. Levine JM, Scrivani PV, Divers TJ, et al. Multicenter case-control study of signalment, diagnostic features, and outcome associated with cervical vertebral malformation-malarticulation in horses. J Am Vet Med Assoc 2010;237:812–822.

5. Levine JM, Adam E, MacKay RJ, et al. Confirmed and presumptive cervical vertebral compressive myelopathy in older horses: A retrospective study (1992–2004). J Vet Intern Med 2007;21:812–819.

6. Moore BR, Reed SM, Biller DS, et al. Assessment of vertebral canal diameter and bony malformations of the cervical part of the spine in horses with cervical stenotic myelopathy. Am J Vet Res 1994;55:5–13.

7. Mayhew IG, Donawick WJ, Green SL, et al. Diagnosis and prediction of cervical vertebral malformation in thoroughbred foals based on semi-quantitative radiographic indicators. Equine Vet J 1993;25:435–440.

8. Papageorges M, Gavin PR, Sande RD, et al. Radiographic and myelographic examination of the cervical vertebral column in 306 ataxic horses. Vet Radiol 1987;28:53–59.

9. Hahn CN, Handel I, Green SL, et al. Assessment of the utility of using intra- and intervertebral minimum sagittal diameter ratios in the diagnosis of cervical vertebral malformation in horses. Vet Radiol Ultrasound 2008;49:1–6.

10. Scrivani PV, Levine JM, Holmes NL, et al. Observer agreement study of cervical-vertebral ratios in horses. Equine Vet J 2011;43:399–403.

11. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307–310.

12. Conover GL, Hildebolt CF, YokoyamaCrothers N. Comparison of linear measurements made from storage phosphor and dental radiographs. Dentomaxillofac Radiol 1996;25:268–273.

13. Cederberg RA, Tidwell E, Frederiksen NL, et al. Endodontic working length assessment. Comparison of storage phosphor digital imaging and radiographic film. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 1998;85:325–328.

14. Hudson NPH, Mayhew IG. Radiographic and myelographic assessment of the equine cervical vertebral column and spinal cord. Equine Vet Educ 2005;17:34–38.

15. Swee RG, Gray JE, Beabout JW, et al. Screen-film versus computed radiography imaging of the hand: A direct comparison. AJR Am J Roentgenol 1997;168:539–542.

16. Kottamasu SR, Kuhns LR, Stringer DA. Pediatric musculoskeletal computed radiography. Pediatr Radiol 1997;27:563–575.

17. Shaw CC, Wang T, King JL, et al. Computed radiography versus screen-film mammography in detection of simulated microcalcifications: A receiver operating characteristic study based on phantom images. Acad Radiol 1998;5:173–180.

18. Weatherburn GC, Ridout D, Strickland NH, et al. A comparison of conventional film, CR hard copy and PACS soft copy images of the chest: Analyses of ROC curves and inter-observer agreement. Eur J Radiol 2003;47:206–214.

19. Alexander K, Joly H, Blond L, et al. A comparison of computed tomography, computed radiography, and film-screen radiography for the detection of canine pulmonary nodules. Vet Radiol Ultrasound 2012;53:258–265.

20. Marolf A, Blaik M, Ackerman N, et al. Comparison of computed radiography and conventional radiography in detection of small volume pneumoperitoneum. Vet Radiol Ultrasound 2008;49:227–232.

21. Wilson AJ, Mann FA, West OC, et al. Evaluation of the injured cervical spine: Comparison of conventional and storage phosphor radiography with a hybrid cassette. Radiology 1994;193:419–422.

22. Mayhew IG, Green SL. Accuracy of diagnosing CVM from radiographs. In: Proceedings of the 39th Annual Congress of the British Equine Veterinary Association. Birmingham: Equine Veterinary Journal Ltd; 2000:74–75.

23. Brage ME, Bennett CR, Whitehurst JB, et al. Observer reliability in ankle radiographic measurements. Foot Ankle Int 1997;18:324–329.

24. Ornetti P, Maillefert JF, Paternotte S, et al. Influence of the experience of the reader on reliability of joint space width measurement. A cross-sectional multiple reading study in hip osteoarthritis. Joint Bone Spine 2011;78:499–505.

25. Bolte H, Jahnke T, Schafer FK, et al. Interobserver-variability of lung nodule volumetry considering different segmentation algorithms and observer training levels. Eur J Radiol 2007;64:285–295.

26. Schafer CB, Sokiranski R, Strayle M, et al. [The value of the supine chest x-ray with digital luminescence radiography in relation to the experience of the observer. A ROC analysis in CT-validated cases]. RoFo 1994;161:25–30.