Received: 19 July 2022 | Revised: 17 April 2023 | Accepted: 24 April 2023
DOI: 10.1111/vsu.13971

CLINICAL RESEARCH

WILEY

Reliability of fluoroscopic examination of nasopharyngeal dorsoventral dimension change in pugs and French bulldogs

Jakub Vodnarek MVDr[1] | **Yasamin Vali DrMedVet, DVM, DVSc**[2] | **Gilles Dupré ProfDrMedVet, DECVS**[1] | **Marlies Dolezal DrNatTech, MSc**[3] | **Manolis Lyrakis PhD**[3] | **Eberhard Ludewig ProfDrMedVetHabil, DECVDI**[2]

[1]Small Animal Surgery, Department for Companion Animals and Horses, University of Veterinary Medicine Vienna, Vienna, Austria

[2]Diagnostic Imaging, Department for Companion Animals and Horses, University of Veterinary Medicine Vienna, Vienna, Austria

[3]Platform for Bioinformatics and Biostatistics, Department of Biomedical Sciences, University of Veterinary Medicine Vienna, Vienna, Austria

Correspondence
Jakub Vodnarek, Small Animal Surgery, Department for Companion Animals and Horses, University of Veterinary Medicine Vienna, Vienna, Austria.
Email: jakub.vodnarek@vetmeduni.ac.at

Abstract

Objective: To compare intra- and interobserver agreements in two-dimensional measurements of changes in nasopharyngeal dimensions during breathing in pugs and French bulldogs.

Study design: Experimental randomized study.

Animals: A total of 20 French bulldogs and 16 pugs.

Methods: Four observers with different levels of experience measured the dorsoventral dimensions of the nasopharynx during inspiration and expiration on fluoroscopy videos. Measurements were performed at the maximal narrowing of the nasopharynx for the functional method and at the level of the tip of the epiglottis for the anatomically adjusted method. The intra- and interobserver agreements of the measurements, ratio of the dynamic nasopharyngeal change (ΔL), and grade of nasopharyngeal (NP) collapse (no, partial or complete) were evaluated.

Results: The functional method resulted in intraobserver correlation coefficients of 0.532 ($p < .01$) and 0.751 ($p < .01$) and interobserver correlation coefficients of 0.378 ($p < .01$) and 0.621 ($p < .01$) for NP collapse grade and ΔL, respectively. The anatomically adjusted method, 0.491 ($p < .01$) and 0.576 ($p < .01$) and 0.495 ($p < .01$) and 0.729 ($p < .01$) for NP collapse grade and ΔL, respectively, were being used. One observer (radiologist) achieved intraobserver correlation coefficients >0.9 for both methods.

Conclusion: Fair interobserver agreement was found for NP collapse grade (functional method), moderate intra- and interobserver agreements were found for NP collapse grade and ΔL (both methods) while intraobserver agreement for ΔL was good (functional method).

Clinical significance: Both methods seem repeatable and reproducible but only for experienced radiologists. The use of ΔL may offer higher repeatability and reproducibility than grade of NP collapse regardless of the method used.

1 | INTRODUCTION

The nasopharynx is the nasal portion of the pharynx that extends from the choanae to the interpharyngeal ostium. The rostral part of the nasopharynx is bounded by solid structures such as the hard palate ventrally, vomer dorsally, and palatine bones on both sides. The middle and caudal portions are bounded dorsally by muscular constrictors (hyopharyngeus, thyropharyngeus, cricopharyngeus, and palatopharyngeus) and dilators (stylopharyngeus) of the pharynx, and the ventral boundary is the mobile soft palate. The soft palate is physiologically able to obliterate the caudal part of the nasopharynx during swallowing via the pressure of the swallowed material and the root of the tongue forcing the soft palate dorsally.[1] Dynamic nasopharyngeal collapse is defined as the partial or complete pathological obliteration of the pharynx due to the dorsal displacement of the soft palate and/or ventral deviation of the pharyngeal wall during inspiration[2,3] (Figure 1). Nasopharyngeal dynamic collapse is considered to be a part of or an important contributor to brachycephalic syndrome.[2] In a recent study, the prevalence of dynamic pharyngeal collapse in brachycephalic dogs undergoing fluoroscopy was 72%.[2] The diagnosis and grading of the severity of nasopharyngeal collapse were based on the visual estimation of nasopharyngeal lumen obliteration using fluoroscopy, which was defined as complete when a complete loss of lumen was observed and partial if the lumen diameter decreased by >50%.[2,3] To the best of our knowledge, no study has compared the intra- and interobserver variability of such estimations. However, low variability is an important prerequisite for the objective detection of nasopharyngeal collapse as a contributing factor to the patient's brachycephalic airway obstruction syndrome and for evaluation of the results of surgical interventions. An ideal method to objectively characterize nasopharyngeal collapse using fluoroscopy is lacking. Measurements can focus on the difference between the maximal and minimal dorsoventral dimensions during one breathing cycle. However, the maximal dorsoventral dimension can appear in another part of the nasopharynx compared to the minimal dimension. Therefore, we have proposed a functional method to measure changes in the dorsoventral dimensions by defining the points for both the maximal and minimal measurement using the location of the minimal dorsoventral dimension of the nasopharynx and comparing the relative dorsoventral change at this location. Furthermore, we have proposed an anatomically adjusted method by defining the maximal and minimal measurement points according to the most rostral extent of the epiglottis.

The objective of our study was to compare the intra- and interobserver agreements in assessing nasopharyngeal collapse in a population of two brachycephalic breeds presented for BOAS at our institution by evaluating the intra- and interobserver variability in measuring changes in the dorsoventral dimensions of the nasopharynx using both the functional and anatomically adjusted methods. Our second aim was to evaluate the clinical applicability of the two methods by comparing the agreement achieved using the respective method.

We hypothesized that these techniques would offer high intra- and interobserver agreement regardless of the breed and the observer's specialty and expertise and would therefore be reliable for evaluating nasopharyngeal collapse. We further hypothesized that the anatomically adjusted method would offer higher intra- and interobserver agreements.

2 | MATERIALS AND METHODS

2.1 | Study design

Anonymized videofluoroscopic examinations of the upper airways of pugs and French bulldogs were retrospectively performed by four observers with different levels of experience (Observer 1: diplomate of the European College of Veterinary Diagnostic Imaging; Observer 2: diplomate of the European College of Veterinary Surgeons; Observer 3: surgery intern; Observer 4: resident of the European College of Veterinary Diagnostic Imaging).

2.2 | Material

The picture archiving and communication system (PACS) at our institution (JiveX, Visus, Essen, Germany) was searched for videofluoroscopic examinations of client-owned pugs and French bulldogs presenting with symptoms of BOAS between January 2014 and January

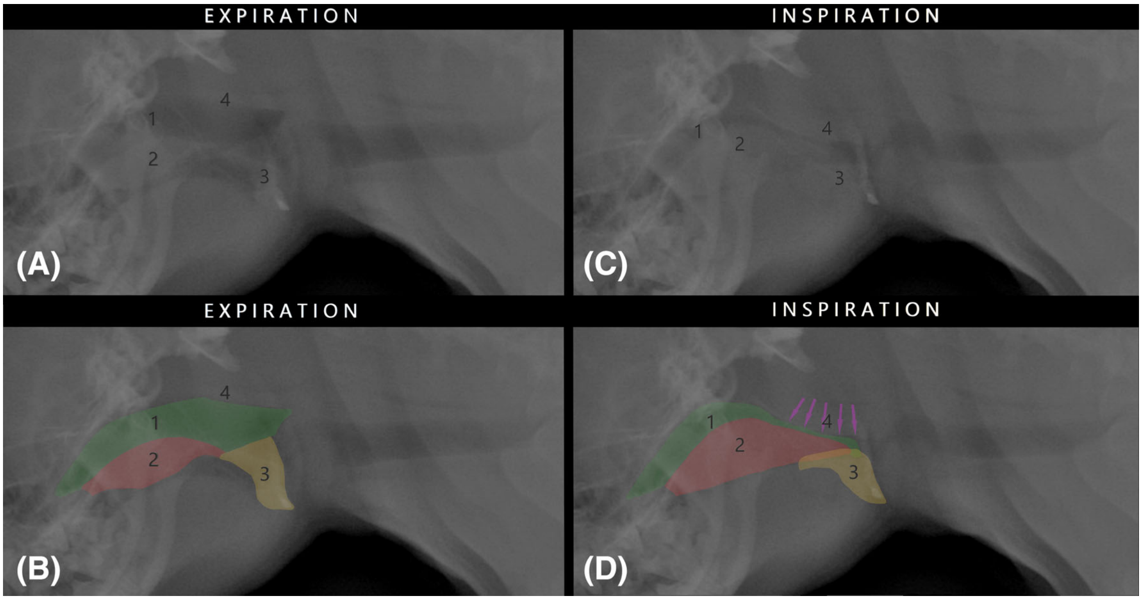

FIGURE 1 Fluoroscopic images of the upper airway of a 1.5-year-old male neutered pug. The images were obtained in awake lateral recumbency. (1) nasopharyngeal air column (shaded green in C+ D), (2) soft palate (shaded pink in C+ D), (3) epiglottis (shaded yellow in C+ D). (4) dorsal nasopharyngeal wall. Note the narrowing of the nasopharyngeal air column during inspiration caused by the dorsal elevation of the soft palate with simultaneous ventral deviation (collapse) of the dorsal nasopharyngeal wall (purple arrows).

2020. Videofluoroscopic examinations were performed as part of the diagnostic protocol for brachycephalic airway syndrome established at our institution.

2.3 | Inclusion and exclusion criteria

For inclusion, the videofluoroscopic examination had to record the nasopharynx for a minimum of two respiration cycles. Poor quality video examinations (e.g., due to inadequate positioning, not recording at least two respiration cycles due to the intolerance of patients to physical restraint, or because of swallowing or panting) were excluded.

2.4 | Fluoroscopic examination

All fluoroscopic examinations were performed using a remote-controlled X-ray diagnostic system with a fluoroscopy table (Axiom Iconos R200, Siemens AG, Erlangen, Germany) and an X-ray tube current of 200 mA and a voltage of 81 kV in the pulsed radiation mode, registering six frames per second. Patients were placed in right lateral recumbency using manual restraint only, paying attention to their tolerance to restraint. Tracheal manipulation and compression were not performed. The examination was performed in the laterolateral view. To obtain valuable information, the total exposure time was determined by a radiology technician.

2.5 | Fluoroscopic recordings

All fluoroscopic evaluations were performed using the PACS at our institution. The recordings fulfilling the inclusion and exclusion criteria were subsequently cut into 9–10-s runs, focusing on the nasopharynx while including at least two respiratory cycles.

The studies were anonymized, exported as DICOM files, and duplicated. The paired recordings were coded and distributed to the observers in a random order using a random number generator function via software available under the GNU License (LibreOffice Calc, The Document Foundation, Berlin, Germany). Each observer received the same set of fluoroscopic studies and performed the measurements using their own laptop screen with the same version of the DICOM viewer. Owing to the duplication and randomization of the videos, each observer performed two measurements for each original video using both methods without knowing whether and when they had previously evaluated the video. The measurements were performed one month after randomization to limit the recall bias of Observer 3, who edited and randomized the videos.

The observers received a brief video tutorial training that explained the functions of the DICOM viewer used

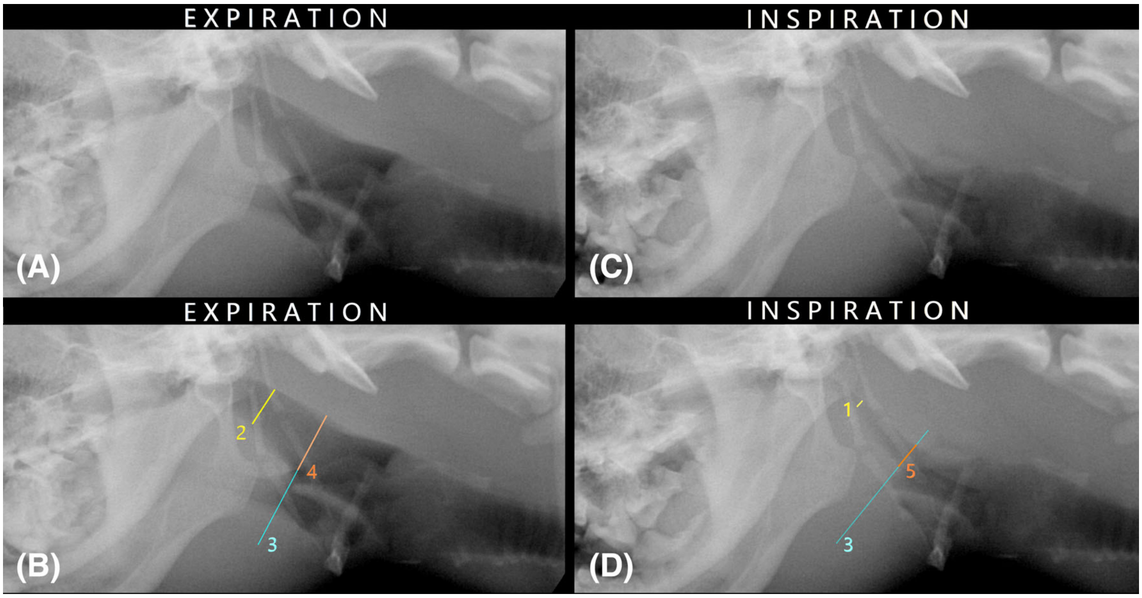

FIGURE 2 Fluoroscopic images of the upper airway of a 2.5-year-old male neutered French bulldog. The images were obtained in awake lateral recumbency. Both the B and D images are copies of their counterparts (A + C), where the measurement lines have been drawn to explain the steps of the measurement: (1) minimum height of the nasopharyngeal air column during inspiration ($L_{Min(FUNCT)}$), (2) maximum height of the nasopharyngeal air column during expiration ($L_{Max(FUNCT)}$), (3) level of the most rostral extent of the epiglottis, (4) The maximum height of the nasopharyngeal air column ($L_{Max(ANAT)}$), (5) minimum height of the nasopharyngeal air column during inspiration ($L_{Min(ANAT)}$).

in the current study (Osirix_Lite software, version 12.x, Pixmeo SARL, Switzerland), described the anatomical boundaries of the nasopharynx as seen on fluoroscopy, defined nasopharyngeal collapse, and explained both measurement methods (see Figure 2).

2.6 | Measurement methods

2.6.1 | Functional method

The functional measurement method consists of the following steps:

1. In videofluoroscopic examination of the nasopharynx with the dog in lateral recumbency, inspiration is defined as the phase of the respiration cycle when the soft palate moves caudodorsally. Conversely, the soft palate deflects rostroventrally during the expiration period. Observers were asked to identify the breathing cycle with the most severe dorsoventral narrowing of the nasopharyngeal lumen.
2. The height of the narrowest nasopharyngeal lumen achieved during inspiration from the chosen breathing cycle was measured ($L_{Min(FUNCT)}$). The measurements were performed perpendicular to the longitudinal axis of the nasopharynx. In some instances, the epiglottis may lift the soft palate dorsally. Therefore, no part of the nasopharynx caudal to the most rostral extremity of the epiglottis was considered in the measurements. Notably, similar to the actual measurement performed perpendicular to the long axis of the nasopharynx, the caudal boundary of the rostral end of the epiglottis was also considered perpendicular to the longitudinal axis of the nasopharynx (blue line in Figure 2).
3. The height of the maximal dimension of the nasopharyngeal lumen achieved throughout the previous or following expiration ($L_{Max(FUNCT)}$) at the same anatomical location as in the previous step was measured (yellow line in Figure 2).

2.6.2 | Anatomically adjusted method

1. Inspiration and expiration of the chosen breathing cycle were identified in the same manner as in the functional method.
2. A tangential line to the rostral-most end of the epiglottic cartilage (blue line in Figure 2) was placed perpendicular to the long axis of the nasopharynx. The minimal ($L_{Min(ANAT)}$) and maximal ($L_{Max(ANAT)}$) heights of the lumen were measured (orange line in Figure 2) alongside the previously created tangent line.

The caliper tool of the DICOM viewer, which was used to perform the linear measurements, allowed measurements with an accuracy of up to two decimal points per millimeter. However, the spatial resolution of the fluoroscopic units was limited. Therefore, the recorded measurements were rounded to the nearest whole millimeter (i.e., $\leq$0.49 to 0, $\geq$0.50 to 1, $\leq$1.49 to 1, $\geq$1.50 to 2, etc.).

The observers recorded paired measurements of minimal and maximal nasopharyngeal dimensions for each fluoroscopic video using both methods (L_{funct}Max, L_{funct}Min, L_{anat}Max, and L_{anat}Min). The randomization key was subsequently revealed, and the observed measurements were assigned back to the patients, distinguishing the first and second measurements of each observer and the measurement method employed for future statistical analyses.

The ratios of dynamic nasopharyngeal changes for each pair of minimal and maximal measurements were calculated using the following formula for both measurement methods:

$$\Delta L = (L_{Max} - L_{Min})/L_{Max}$$

A nasopharyngeal collapse grade was then assigned according to a previously published three-tier grading (no collapse: $\Delta L < 0.5$, partial collapse: $\Delta L \geq 0.5$ and <1, and complete collapse: $\Delta L = 1$).[2,3]

Using the anonymization and randomization key, a dataset of paired measurements for the first and second attempts and the respective ratios of the dynamic nasopharyngeal change for each animal and observer was created and prepared for statistical evaluation.

2.7 | Statistical analysis

Using the icc function (R package irr, R version 4.0.2),[4,5] the paired measurements of the minimal and maximal nasopharyngeal dimensions of both methods (L_{Max}, L_{Min}) and paired ratios of the dynamic nasopharyngeal change (ΔL) were compared for the intraobserver agreement for all observers combined (global correlation coefficient) and each observer separately. The means of the paired measurements and ratios were compared for interobserver variability across all observers (global correlation coefficient) and for each pair of observers separately. The Bonferroni-Holm method was used for multiple testing corrections. Similarly, the assigned grade of nasopharyngeal collapse was analyzed using the function kappam.fleiss (R package irr). Statistical significance was set at an alpha cutoff of 5% after multiple testing corrections.

The reliability of the observed intraclass correlation coefficient (ICC) for intra- and interobserver agreement in the measurement of L_{Max}, L_{Min}, and ΔL values was interpreted based on previously published guidelines, where ICC values <0.5 indicate poor reliability, values between 0.5 and 0.75 indicate moderate reliability, values between 0.75 and 0.9 indicate good reliability, and values greater than 0.9 indicate excellent reliability.[6]

The strength of the intra- and interobserver agreement for assigning the grade of nasopharyngeal collapse was interpreted based on previous recommendations that considered the kappa statistic.[7] Notably, κ values <0.20 were considered poor, 0.21–0.40 were considered fair, 0.41–0.60 were considered moderate, 0.61–0.80 were considered good, and 0.81–1.00 were considered very good.

3 | RESULTS

A total of 43 fluoroscopic videos of the upper airways of French bulldogs and 35 videos of the upper airways of pugs were obtained from the PACS at our institution. However, videos of only 20 French bulldogs and 16 pugs fulfilled the inclusion criteria.

3.1 | Intraobserver variability for L_{Max}

The global correlation coefficient for intraobserver variability for the measurement of L_{Max} was 0.878 ($p < .01$) for the functional method and 0.785 ($p < .01$) for the anatomically adjusted method and was therefore interpreted as being good. Observer 1 achieved excellent, and the highest, consistency between the first and second measurements for both the functional (0.972, $p < .01$) and anatomically adjusted methods (0.973, $p < .01$) (Table 1).

3.2 | Interobserver variability for L_{Max}

The global correlation coefficient for interobserver variability for the measurement of L_{Max} was 0.857 ($p < .01$) for the functional method and 0.763 ($p < .01$) for the anatomically adjusted method and was therefore interpreted as being good (Table 2).

3.3 | Intraobserver variability for L_{Min}

The global correlation coefficient for intraobserver variability for the measurement of L_{Min} was 0.795 ($p < .01$)

TABLE 1 Intraobserver variability for L_{Max}.

	Functional method		Anatomically adjusted method	
Observer	Correlation coefficient	*p*-value	Correlation coefficient	*p*-value
Global	0.878	<.0001	0.785	<.0001
1	0.972	<.0001	0.973	<.0001
2	0.814	<.0001	0.737	<.0001
3	0.870	<.0001	0.842	<.0001
4	0.865	<.0001	0.627	<.0001

TABLE 2 Interobserver variability for L_{Max}.

	Functional method		Anatomically adjusted method	
Observer	Correlation coefficient	*p*-value	Correlation coefficient	*p*-value
Global	0.857	<.0001	0.763	<.0001
1 vs. 2	0.904	<.0001	0.724	<.0001
1 vs. 3	0.830	<.0001	0.771	<.0001
1 vs. 4	0.865	<.0001	0.670	<.0001
2 vs. 3	0.840	<.0001	0.891	<.0001
2 vs. 4	0.830	<.0001	0.747	<.0001
3 vs. 4	0.871	<.0001	0.735	<.0001

TABLE 3 Intraobserver variability for L_{Min}.

	Functional method		Anatomically adjusted method	
Observer	Correlation coefficient	*p*-value	Correlation coefficient	*p*-value
Global	0.795	<.0001	0.676	<.0001
1	0.949	<.0001	0.961	<.0001
2	0.735	<.0001	0.452	.00228
3	0.716	<.0001	0.702	<.0001
4	0.699	<.0001	0.676	.000101

TABLE 4 Interobserver variability for L_{Min}.

	Functional method		Anatomically adjusted method	
Observer	Correlation coefficient	*p*-value	Correlation coefficient	*p*-value
Global	0.700	<.0001	0.766	<.0001
1 vs. 2	0.834	<.0001	0.655	<.0001
1 vs. 3	0.532	.000208	0.816	<.0001
1 vs. 4	0.706	<.0001	0.825	<.0001
2 vs. 3	0.606	<.0001	0.699	<.0001
2 vs. 4	0.816	<.0001	0.683	<.0001
3 vs. 4	0.666	<.0001	0.861	<.0001

(good) for the functional method and 0.676 ($p < .01$) (moderate) for the anatomically adjusted method. Observer 1 achieved excellent consistency between the first and second measurements for both the functional (0.949, $p < .01$) and anatomically adjusted methods (0.961, $p < .01$) (Table 3).

3.4 | Interobserver variability for L_{Min}

The global correlation coefficient for the interobserver variability for the measurement of L_{Min} was 0.7 ($p < .01$) (moderate) for the functional method and 0.766 ($p < .01$) (good) for the anatomically adjusted method (Table 4).

Observer	Functional method: Correlation coefficient	Functional method: p-value	Anatomically adjusted method: Correlation coefficient	Anatomically adjusted method: p-value
Global	0.751	<.0001	0.576	<.0001
1	0.921	<.0001	0.940	<.0001
2	0.611	<.0001	0.445	.00533
3	0.676	<.0001	0.558	.000434
4	0.706	<.0001	0.361	.013

TABLE 5 Intraobserver variability for the ratio of the dynamic change in nasopharyngeal dimensions.

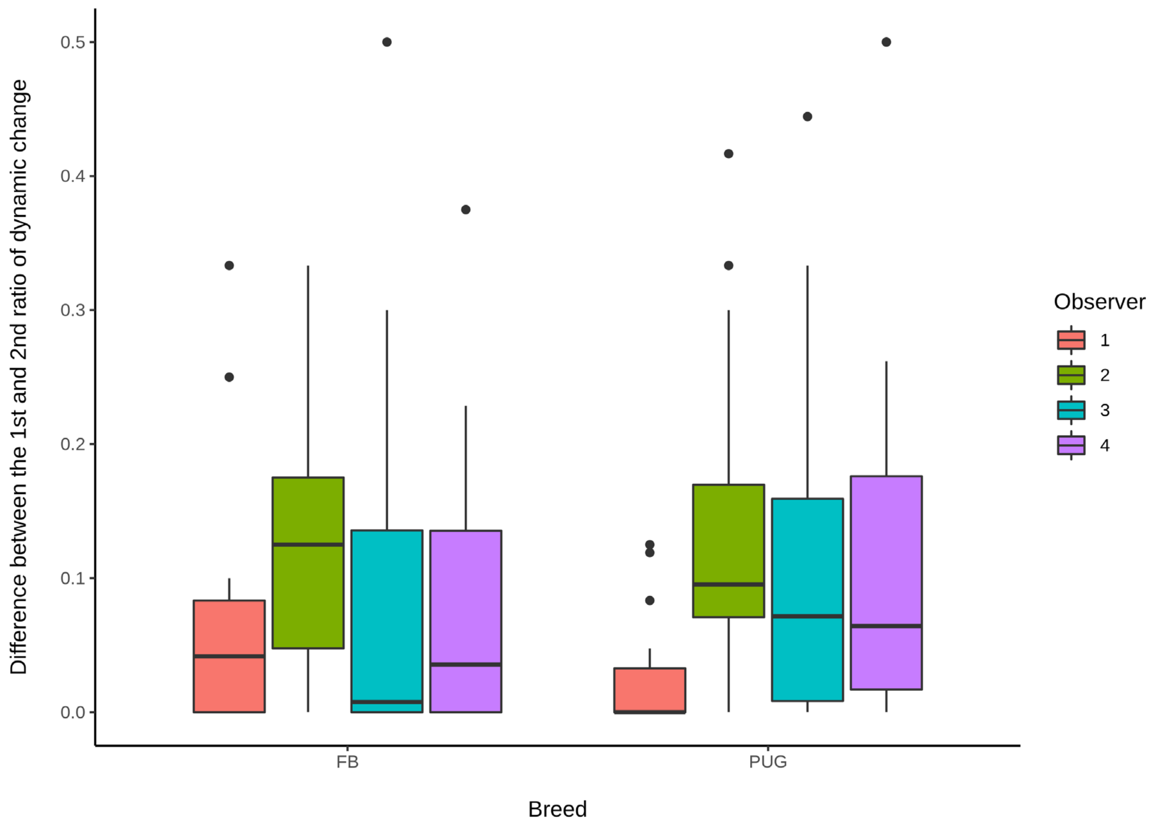

FIGURE 3 Boxplot of intraobserver variability in measuring the ratio of the nasopharyngeal dynamic change using the functional method. The boxes represent the 25th to 75th interquartile range (IQR) of intraobserver differences in the ratio of the dynamic nasopharyngeal change. The transverse line through the boxes represents the median. The upper and lower whiskers (vertical black lines) represent the 75th percentile +1.5 * IQR and 25th percentile −1.5 * IQR, respectively. The dots represent the outliers. The y-axis has a step size of 0.1 (10% difference in the ratio of the dynamic nasopharyngeal change). Observer (1) diplomate ECVDI, Observer (2) diplomate ECVS, Observer (3) surgery intern, Observer (4) resident ECVDI. Observer 1 achieved the most consistent measurements among all the observers and performed better in pugs than in French bulldogs (FB).

3.5 | Intraobserver variability for ΔL

The global correlation coefficient for intraobserver variability for ΔL was 0.751 ($p < .01$) (good) for the functional method and 0.576 ($p < .01$) (moderate) for the anatomically adjusted method. Observer 1 achieved excellent, and the highest, consistency between the first and second measurements for both the functional (0.921, $p < 0.01$) and anatomically adjusted methods (0.94, $p < 0.01$) (Table 5). The intraobserver variability for ΔL is plotted in Figure 3 for the anatomical method and in Figure 4 for the functional method.

3.6 | Interobserver agreement for ΔL

The global correlation coefficient for interobserver agreement for the measurement of ΔL was 0.621 ($p < .01$) (moderate) for the functional method and 0.729 ($p < .01$)

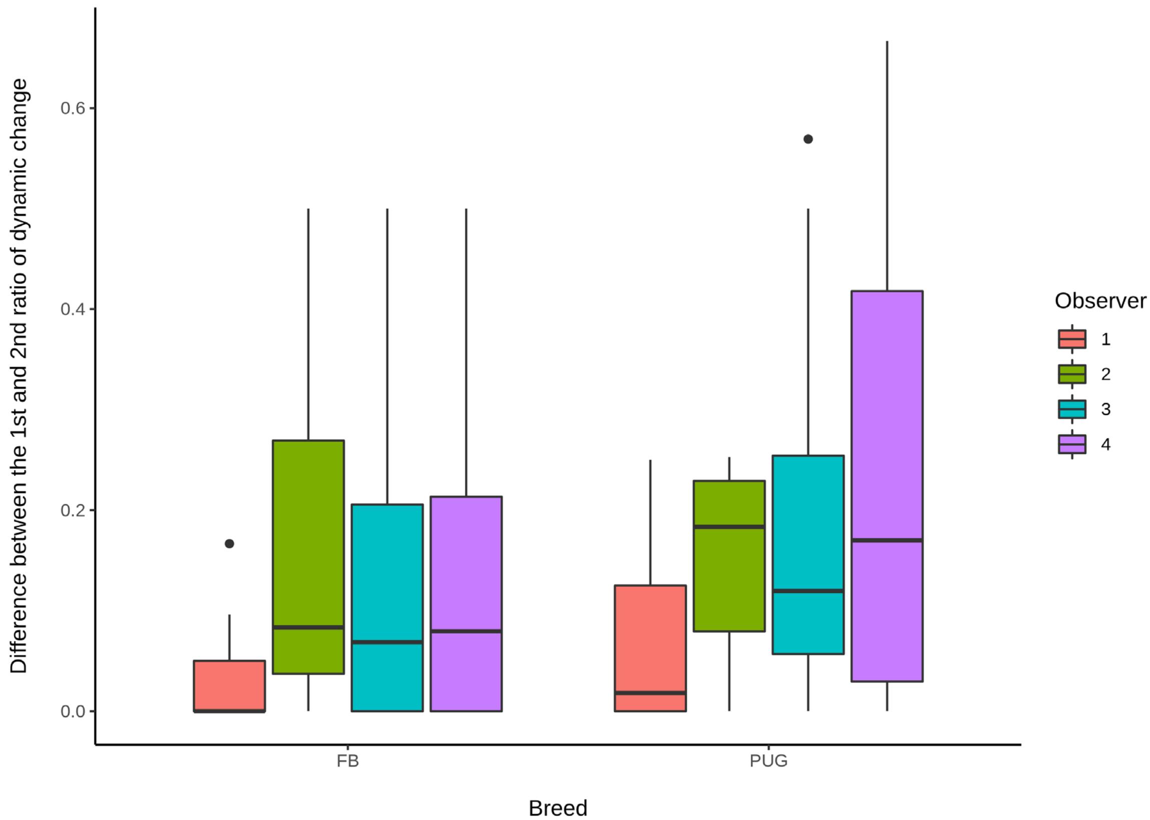

FIGURE 4 Boxplot of intraobserver variability in measuring the ratio of the nasopharyngeal dynamic change using the functional method. The boxes represent the 25th to 75th interquartile range (IQR) of intraobserver differences in the ratio of the dynamic nasopharyngeal change. The transverse line through the boxes represents the median. The upper and lower whiskers (vertical black lines) represent the 75th percentile +1.5 * IQR and 25th percentile −1.5 * IQR, respectively. The dots represent the outliers beyond this. The step size of the y-axis is 0.1 (10% difference in the ratio of the dynamic nasopharyngeal change). Observer (1) diplomate ECVDI, Observer (2) diplomate ECVS, Observer (3) surgery intern, Observer (4) resident ECVDI. Observer 1 achieved the most consistent measurements among all the observers and performed better in French bulldogs (FB) than in pugs.

TABLE 6 Interobserver variability for the ratio of the dynamic change in nasopharyngeal dimensions.

	Functional method		Anatomically adjusted method	
Observer	**Correlation coefficient**	***p*-value**	**Correlation coefficient**	***p*-value**
Global	0.621	<.0001	0.729	<.0001
1 vs. 2	0.812	<.0001	0.709	<.0001
1 vs. 3	0.437	.00314	0.795	<.0001
1 vs. 4	0.645	<.0001	0.747	<.0001
2 vs. 3	0.523	.0018	0.610	<.0001
2 vs. 4	0.766	<.0001	0.730	<.0001
3 vs. 4	0.514	.0018	0.755	<.0001

(moderate) for the anatomically adjusted method (Table 6).

Of all the measured ΔLs, 234 (81.25%) and 219 (76.04%) values were between 0.6 (including 0.6) and 1 for the functional and anatomically adjusted methods, respectively (Figure 5).

3.7 | Intra- and interobserver agreement for the grade of dynamic nasopharyngeal collapse

The global correlation coefficient for intraobserver agreement was 0.532 ($p < .01$) for the functional method and

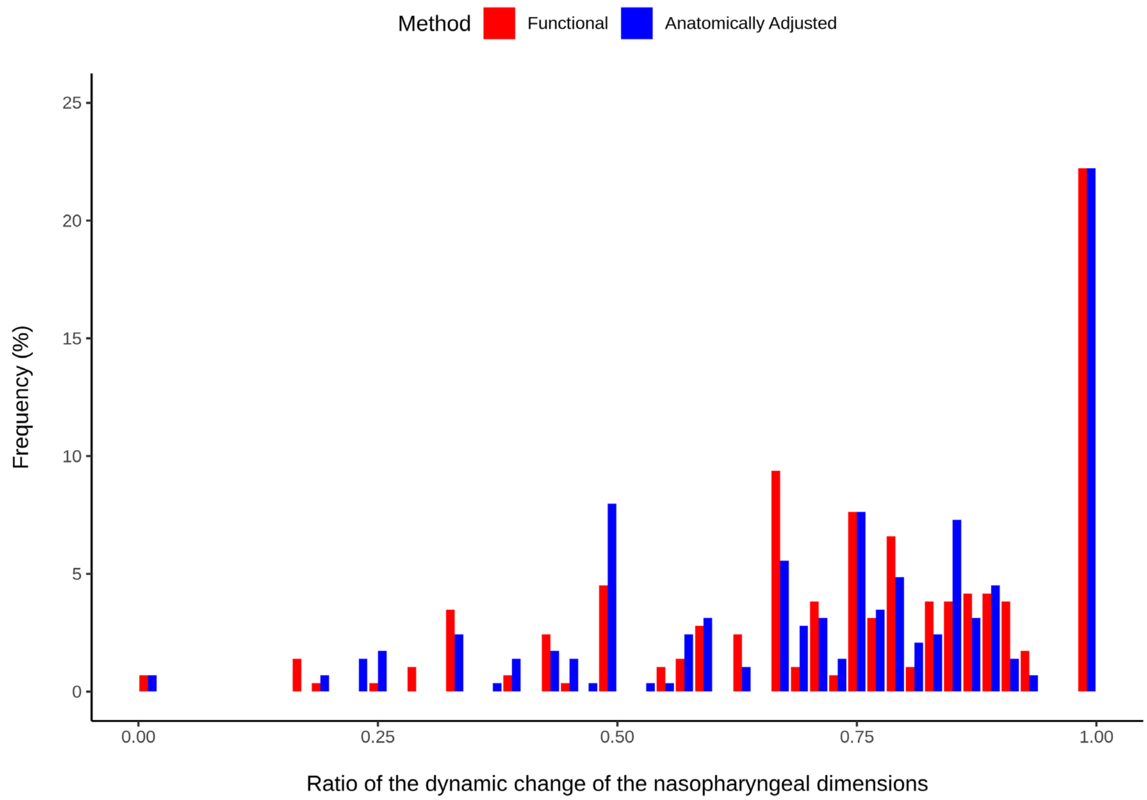

FIGURE 5 Histogram of the dynamic change ratios (pooled across all observers) of the nasopharyngeal dimensions using the functional (red) and anatomically adjusted (blue) methods.

TABLE 7 Intraobserver variability for assigning a grade of dynamic nasopharyngeal collapse.

	Functional method		Anatomically adjusted method	
Observer	**Correlation coefficient (κ)**	***p*-value**	**Correlation coefficient (κ)**	***p*-value**
Global	0.491	<.0001	0.532	<.0001
1	0.887	<.0001	0.803	<.0001
2	0.214	.092	0.179	.000775
3	0.408	.00369	0.499	<.0001
4	0.388	.00412	0.547	.159

TABLE 8 Interobserver variability for assigning a grade of dynamic nasopharyngeal collapse.

	Functional method		Anatomically adjusted method	
Observer	**Correlation coefficient (κ)**	***p*-value**	**Correlation coefficient (κ)**	***p*-value**
Global	0.495	0	0.378	<.0001
1 vs. 2	0.316	.026	0.659	<.0001
1 vs. 3	0.620	<.0001	0.310	.042
1 vs. 4	0.618	<.0001	0.494	.000437
2 vs. 3	0.278	.027	0.211	.146
2 vs. 4	0.666	<.0001	0.397	.00852
3 vs. 4	0.450	.00106	0.231	.146

0.491 ($p < .01$) for the anatomically adjusted method; therefore, it was interpreted as being moderate for both methods. Observer 1 achieved very good intraobserver agreement ($\kappa = 0.887$; $p < .01$) for the anatomical method and good intraobserver agreement ($\kappa = 0.803$; $p < .01$) for the functional method (Tables 7 and 8).

The global correlation coefficient for interobserver agreement was 0.495 ($p < .01$) (moderate) for the anatomically adjusted method and 0.378 ($p < .01$) (fair) for the functional method.

Each nasopharynx was examined by four observers twice, resulting in 288 diagnosed grades of nasopharyngeal collapse for each measurement method. Considering these observations as a population, no collapse was assigned in 31 (10.76%) and 35 (12.15%) instances, and partial collapse was assigned in 193 (67.01%) and 189 (65.63%) instances using functional and anatomically adjusted methods, respectively. Complete collapse was observed in 64 instances (22.22%) using both the methods.

However, when considering only the mean values from the first and second observations performed by

Observer 1, 36 grades were assigned using both the methods. Of these, no collapse was assigned in six (16.76%) and five (13.88%) instances, partial collapse in 27 (75%) and 23 (63.88%) instances, and complete collapse in three (8.33%) and eight (22.22%) instances using functional and anatomically adjusted methods, respectively.

4 | DISCUSSION

Both methods offered good intra- and interobserver agreement for the measurement of L_{Max}. While the intraobserver agreement for the measurement of L_{Min} was good for the functional method, it was only moderate for the anatomically adjusted method. Conversely, the anatomically adjusted method offered good interobserver agreement for L_{Min}, whereas it was moderate for the functional method. We found good and moderate intraobserver agreement for ΔL using the functional and anatomically adjusted methods, respectively. The interobserver agreement for ΔL using both the functional and anatomically adjusted methods was good. Furthermore, we found moderate intra- and interobserver agreement for the grade of dynamic nasopharyngeal collapse using both methods, except for the interobserver agreement in grading using the anatomically adjusted method, which was fair (0.378).

There are several sources of variability in both the measurement methods. For instance, observers might have difficulty aligning the fiducial marker on the selected point on the screen. In addition, they might have difficulty deciding the point of the minimum height of the nasopharyngeal lumen and recognizing whether there was a complete or near-complete collapse. In particular, attention was needed to distinguish the collapsing nasopharynx during inspiration from physiological swallowing movements. The fluoroscopic videos were edited to include at least two breathing cycles undisturbed by swallowing or other motions. The rationale behind this design was to save the time of the observers, as substantial stoppage during recordings has poor diagnostic value. Observers were asked to choose the breathing cycle with the most severe dorsoventral narrowing of the nasopharynx. Therefore, the observers may have measured different breathing cycles. Theoretically, we could have asked the observers to evaluate a specified breathing cycle. Although the variability of the change in the dorsoventral nasopharyngeal dimensions between different breathing cycles has not yet been reported, it appears reasonable to assume that such variability exists and, therefore, allowing the observers to choose which breathing cycle to measure would be more appropriate for clinical use. Thus, minimal, maximal, and mean changes in dorsoventral nasopharyngeal dimensions could be the subjects of further studies.

Owing to the retrospective nature of the study, we could not choose a position different from lateral recumbency, such as normal standing or sternal recumbency, as the remote-controlled X-ray diagnostic system with a fluoroscopy table (Axiom Iconos R200, Siemens AG, Erlangen, Germany) employed at the time would not allow such positioning. Fixating a brachycephalic dog in the lateral position may increase its stress level, potentially leading to the deterioration of its already compromised breathing. Therefore, the protocol at our institution at the time was to abort the examination if the patient did not tolerate the lateral position.

Despite these limitations, one observer (a diplomate of the European College of Diagnostic Imaging) achieved excellent intraobserver agreement in measuring the ratio of dynamic nasopharyngeal change using both methods (correlation coefficients >0.9). Consequently, the observer reached very good intraobserver agreement for the grading of nasopharyngeal collapse using the functional method and good agreement using the anatomically adjusted method (correlation coefficients 0.887 and 0.803, respectively). It is plausible that the effects of profound training of this observer allowed for higher agreement, although larger groups of observers are needed to more strongly support this conclusion.

Our study was not designed to assess the prevalence of nasopharyngeal collapse. However, when considering only the mean values from the first and second observations performed by Observer 1, partial or complete collapse was widespread (83.3% for the functional method and 86.1% for the anatomically adjusted method). This is slightly higher than the previously reported incidence of 72% for brachycephalic breeds.[7] The reason for this increase in incidence could be explained by the increased sensitivity of our methodology, geographical differences between studied populations, and/or coincidence due to selection bias caused by the exclusion of several fluoroscopic examinations from our study due to poor quality. In our study, the anatomical technique in the hands of the radiologist led to more common identification of complete than partial collapse. This is unexpected because the anatomically adjusted method places the measurement in a given location. The expectation was that because of that, some underestimation of the severity of the collapse could occur. The observations were performed on dogs with brachycephalic airway obstruction syndrome before surgery. An overlong soft palate is expected to occur commonly. In such a situation, the tip of the epiglottis elevates the soft palate, causing further narrowing of the airway beyond what would be caused by the collapsing dorsal structures. Therefore, the observers were instructed not to consider the parts of the nasopharynx caudal to the most rostral extremity of the epiglottis for the measurements. It is, therefore, possible that while Observer 1 was using the functional method, he might have found some complete collapses to be located too caudally for consideration but have found those to be included in the examined area after performing the measurement with the help of the tangential line to the rostral-most end of the epiglottic cartilage using the anatomically adjusted method.

Although using an established grading system might seem easier for clinical communication, introducing any cutoffs to continuous data leads to an increase in the variability of the assessment. For example, dogs with a ΔL of 0.49 and those with a ΔL of 0.51 are likely to be clinically similarly affected despite having different grades. Conversely, dogs with a respective ΔL of 0.51 and 0.99 would likely be differently affected, despite having the same grade. We hypothesized that both methods (functional and anatomically adjusted) would offer high intra- and interobserver agreement and would therefore be reliable for evaluating nasopharyngeal collapse in two brachycephalic breeds (French bulldogs and pugs). An anatomically adjusted method was developed to reduce the variability of the measurements; however, our results did not support this hypothesis because the functional method delivered marginally better agreement for ratios and grading with the exception of the interobserver agreement for grading, where the anatomically adjusted method performed better. Due to the considerable variability among observers, we did not consider a statistical comparison between the two methods. However, such a comparison might be matter of future studies employing only observers with speciality training in imaging.

In conclusion, the global intra- and interobserver agreement of two-dimensional measurements of the changes in nasopharyngeal dimensions during breathing in a population of brachycephalic dogs was good to moderate, indicating considerable variability in fluoroscopic evaluation of dynamic changes in the dorsoventral nasopharyngeal dimensions. Although the repeatability of the proposed methodology among veterinarians without imaging specialty training may be lower, both techniques may achieve higher repeatability among experienced radiologists. None of the methods was superior to the other. Furthermore, we conclude that the use of the ratio of the dynamic change of the dorsoventral nasopharyngeal dimensions may be more appropriate than the use of grades; it not only avoid introducing cutoffs with unknown clinical relevance, but it also offered marginally better global intra- and interobserver agreements in our study. Further studies comparing interobserver agreement among trained specialists are required to determine which of the studied methods offers higher intra- and interobserver repeatability and clinical usefulness. Furthermore, future research should investigate the impact of nasopharyngeal collapse on respiration, prognosis, and the effect of airway surgery on further progression or improvement of nasopharyngeal collapse.

AUTHOR CONTRIBUTIONS

Vodnarek J, MVDr: Contributed to the study design, proposal of the measurement method, and acquisition, randomization, and blinding of the data. The same author also acquired and prepared the data for the observers, performed the blinded observations, collected data from the observers, prepared data for statistical evaluation, drafted the manuscript, and approved the final version for publication. Ludewig E, ProfDrMedVetHabil, DECVDI and Vali Y, DrMedVet, DVM, DVSc: Contributed to the study design, performed the blinded observations, edited the manuscript, and approved the final version for publication. Dupré G, ProfDrMedVet, DECVS: Contributed to the conception of the study and study design, performed the blinded observations, edited the manuscript, and approved the final version for publication. Lyrakis M, PhD and Dolezal M, DrNatTech, MSc: Performed the statistical analyses of all the data, edited the manuscript, and approved the final version for publication.

ACKNOWLEDGMENTS

The authors would like to thank to PD Dr. med. vet. habil. Barbara Bockstahler for the supervision of our research group and help with manuscript editing.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest related to this report.

ORCID

Jakub Vodnarek https://orcid.org/0000-0002-8043-8189
Yasamin Vali https://orcid.org/0000-0002-6090-0663

REFERENCES

1. Pollard RE. Videofluoroscopic evaluation of the pharynx and upper esophageal sphincter in the dog: a systematic review of the literature. *Front Vet Sci*. 2019;6(MAR):1-9.
2. Pollard RE, Johnson LR, Marks SL. The prevalence of dynamic pharyngeal collapse is high in brachycephalic dogs undergoing videofluoroscopy. *Vet Radiol Ultrasound*. 2018;59(5):529-534.
3. Rubin JA, Holt DE, Reetz JA, Clarke DL. Signalment, clinical presentation, concurrent diseases, and diagnostic findings in 28 dogs with dynamic pharyngeal collapse (2008–2013). *J Vet Intern Med*. 2015;29(3):815-821.
4. Gamer M, Lemon J, Fellows I, Singh P. irr: Various Coefficients of Interrater Reliability and Agreement. R package version 0.84.1. The Comprehensive R Archive Network. 2019 https://cran.r-project.org/web/packages/irr/index.html

5. Core Team. R: A language and environment for statistical computing, Vienna Austria. 2020 https://www.R-project.org Accessed December 25, 2021.
6. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155-163.
7. Brennan P, Silman A. Statistical methods for assessing observer variability in clinical measures. *Br Med J*. 1992;304(6840):1491-1494.

How to cite this article: Vodnarek J, Vali Y, Dupré G, Dolezal M, Lyrakis M, Ludewig E. Reliability of fluoroscopic examination of nasopharyngeal dorsoventral dimension change in pugs and French bulldogs. *Veterinary Surgery*. 2024;53(1):84-95. doi:10.1111/vsu.13971