European College of Veterinary Surgeons

Short Abstract Form for Publication

1. Title of abstract: VARIABILITY OF RESTING ENDOSCOPIC GRADING FOR ASSESSMENT OF LARYNGEAL FUNCTION IN HORSES.

2. This abstract is submitted by: Rachel Salz

3. Names of authors: presenting author a) R.O. Salz
other authors b) J.D. Perkins
c) J. Schumacher
d) L. Livesey
e) R. Piercy
f) S. Z. Barakzai

4. Name, complete address, phone and fax number of presenting author:

Name: R.O. Salz

Address: Department of Veterinary Clinical Sciences, Royal Veterinary College, Hawkshead Lane, North Mymms, Hatfield, Hertfordshire AL9 7TA, U.K.

Phone: 01707 666297
Email: rsalz@rvc.ac.uk

Tick **only one: scientific /resident** or poster session

Category:

- ○ **Scientific session – oral**:
  - ☐ • SA General
  - ☐ • SA Orthopaedics
  - ☐ • LA General
  - ☐ • LA Orthopaedics

- ○ **Resident's Forum – oral:**
  - ☐ large animal
  - ☐ small animal

Oral presentation:
- **only Powerpoint from data projector**
- **on CD or USB Key**

☐ **NOT** for poster

- ○ **Poster**
  - ☐ large animal
  - ☐ small animal

Poster format A0
(width 90 / heighth 120 cm)

☐ **Indicate in case you accept also to present your abstract as poster**

Type abstract here, including title, author(s) and institution(s), city, country and (9 or 10 point font: times) Stay within border!

VARIABILITY OF RESTING ENDOSCOPIC GRADING FOR ASSESSMENT OF LARYNGEAL FUNCTION IN HORSES

R.O. Salz, J.D. Perkins, J. Schumacher[†], L. Livesey[†], R. Piercy, S. Z. Barakzai[‡]

Department of Veterinary Clinical Sciences, Royal Veterinary College, North Mymms, Hertfordshire, U.K. [‡] Department of Veterinary Clinical Studies, Faculty of Veterinary Medicine, Royal (Dick) School of Veterinary Studies, Midlothian, U.K. [†]Department of Veterinary Clinical Sciences, Auburn University, Auburn, Alabama, U.S.A.

Introduction: The variability of endoscopic grading of arytenoid cartilage movement is uncertain. The purpose of this study was to determine the observer and horse variability of grading arytenoid cartilage movement in horses during resting endoscopic examination, using a seven-grade system.

Materials and methods: Endoscopic recordings of the upper respiratory tract made at rest in 270 draft horses were reviewed independently by two veterinarians to assess interobserver variability. Laryngeal function was assessed subjectively using a seven-grade system. Grading was repeated by both examiners in 80 randomly selected recordings in order to assess intraobserver variability. In 120 horses, endoscopic recordings were repeated after 24-48 hours, and these were graded by both veterinarians to assess intrahorse variability.

Results: The mean weighted κ statistic for concordance within examiners was 0.867, with a mean intraobserver agreement of 76.3%. The weighted κ statistic for concordance between the two examiners was 0.765, with an interobserver agreement of 63.1%. Of the horses receiving two endoscopic examinations, the same grade was assigned to 41.7% of horses at the second examination. The mean weighted κ statistic for concordance between the grade assigned at first vs. second examinations was 0.588, indicating only moderate agreement.

Discussion/Conclusion: Intraobserver and interobserver reliability of resting endoscopic grading of arytenoid cartilage movement using a seven-grade system was high when examinations were conducted by experienced veterinarians. However there was moderate daily intrahorse variability, suggesting that results of resting endoscopic examinations performed on a single day should be interpreted with caution, particularly when making decisions on the assessment of horses at the time of sale.

Sign. (presenting author): _____ **Sign. (Resident Advisor):** _____

European College of Veterinary Surgeons

Long Abstract Form for Blinded Evaluation

(no names or other identification on this form)

Title of abstract:

Variability of Resting Endoscopic Grading for Assessment of Laryngeal Function in Horses

Introduction: There were two main purposes to this study. The first was to determine the intraobserver and interobserver variability of endoscopic grading of arytenoid cartilage movement in a large number of Clydesdale and Belgian/Percheron horses, using a seven-grade system. The second was to assess intrahorse variability by investigating whether the grade assigned to a particular horse altered when endoscopic examination was repeated after 24-48 hours.

Materials and methods:

Animals

Laryngeal function was assessed in 120 Belgian/Percheron horses and 150 Clydesdale horses using digital videoendoscopy. The Belgian/Percheron and Clydesdale horses were randomly selected from herds in the USA and Scotland respectively. The breed, age and sex of each horse were recorded.

Procedure

Horses were restrained without sedation using a nose twitch. A flexible digital videoendoscope was introduced through the right nostril, advanced to the pharynx and positioned centrally so that the appearance and movements of the larynx could be observed. Laryngeal activity was examined during respiration at rest and also following swallowing, which was induced a minimum of 3 times in each horse. Each examination was digitally recorded and stored for subsequent review. Each recording lasted 1-4 minutes depending on the cooperation of the horse. Endoscopic examinations were repeated in 120 Belgian/Percheron horses after 24-48 hours.

Horses with upper respiratory tract abnormalities other than RLN, such as severe pharyngeal lymphoid hyperplasia or arytenoid chondritis, were discarded from the study. In addition, incomplete recordings were discarded (for example, if swallowing was not successfully induced 3 times or if full abduction of the right arytenoid cartilage was not achieved). The recordings were given to an independent reviewer to be renumbered and reordered. These were then independently graded by 2 veterinarians (observer 1 and observer 2) experienced with endoscopic examination of horses, who subjectively assessed arytenoid cartilage movements using a four-grade system with sub-grades (Havemeyer grading system, Dixon *et al.* 2003, Table 1). Grading was repeated by both veterinarians on 80 randomly chosen recordings, which had been assigned further coded numbers by the independent reviewer, at a minimum of 6 months after initial grading.

Statistical analysis

Agreement on grading of laryngeal function within and between observers was summarised as frequency of agreement and range of disagreement. Agreement between the grades assigned at first vs. second examinations in those horses receiving 2 endoscopic examinations on different days was assessed in the same way. To assess intraobserver variability, mean percentage agreements for first vs. second grading by each observer were calculated within each grade. This was done by calculating the mean of the percentage agreement within each grade when first grading was compared to second grading, and when second grading was compared to first grading. Mean percentage agreements within each grade for interobserver variability were calculated similarly, by calculating the mean of the percentage agreement when observer 1 was compared to observer 2, and when observer 2 was compared to observer 1.

Weighted κ statistics, reported with standard errors, were calculated for all pairs of observations (pairs of examinations within and between observers and first vs. second examination for each horse). Weighted κ statistics take into account the magnitude of the difference between observers for a particular observation. Disagreement is considered more significant if it is by more than 1 grade. The κ statistics were considered to indicate poor ($\kappa<0.20$), fair ($0.20\leq\kappa\leq0.40$), moderate ($0.41\leq\kappa\leq0.60$), good ($0.61\leq\kappa\leq0.80$) and excellent ($\kappa>0.80$) agreement. The differences in grades between first vs. second grading for each observer, between the 2 observers and between first vs. second grading for horses receiving 2 endoscopic examinations were not found to be normally distributed (using the Kolmogorov-Smirnov test for normal distribution), therefore, statistical analysis was performed using non-parametric Kruskall-Wallis tests (Bonferroni correction was applied for multiple comparisons and significance was accepted at $P<0.0024$ as indicating a significant difference between grades). Differences between individual pairs of grades for all pairs of observations were tested using Mann-Whitney U tests (significance was accepted at $P<0.0024$). For the purpose of statistical analysis, the grades were re-numbered 1-7 and a linear increase between grades was assumed.

Results

Sixteen of the Belgian/Percheron horses and 21 of the Clydesdale horses were discarded due to other upper respiratory tract abnormalities or incomplete recordings. This left a population of 233 horses for assessment of interobserver variability and 104 horses for assessment of intrahorse variability. In the group of Belgian/Percheron horses, there were 82 Belgian and 22 Percheron horses. Overall, there were 160 mares, 58 geldings and 15 stallions. The mean age of the horses was 7.5 years (range 1-24 years).

Intraobserver variability

Both observers assigned the same grade on second viewing of the same endoscopic recording (intraobserver agreement) in 76.3% of recordings. Of the 80 recordings graded twice by each observer, there were 50 recordings (62.5%) in which both observers assigned the same grade, and 8 recordings (10%) in which both observers assigned different grades. All of the 8 horses in which both observers assigned different grades were graded 2.1, 2.2 or 3.1 at 1 of the viewings, by each observer. The weighted κ statistics for concordance within observers was 0.870 (standard error 0.071) for observer 1 and 0.863 (standard error 0.073) for observer 2, both of which correlate to excellent agreement.

Interobserver variability

Both observers assigned the same grade to 63.1% of horses. The observers differed by 1 grade in 32.2% of horses and differed by 2 grades in 4.7% of horses. The weighted κ statistic for concordance between observers was 0.765 (standard error 0.040), which correlates to good agreement. The 2 observers were found to disagree significantly more often over a grade 2.1, when compared to any other grade ($P<0.0024$ with Bonferroni correction applied).

Intrahorse variability

Of the 104 horses that received 2 endoscopic examinations, fewer than half the horses (41.7%) were graded the same at the second examination (mean percentage agreement of the 2 observers). 42.3% of horses were given a grade which differed by 1, with 16.6% receiving an improved grade and 25.7% receiving a worse grade. 13.2% of horses were given a grade which differed by 2, with 4.9% receiving an improved grade and 8.4% receiving a worse grade. 2.3% of horses were given a grade which differed by 3, with 1.3% receiving an improved grade and 1.1% receiving a worse grade. One horse improved by 4 grades. The weighted κ statistics for concordance between grade assigned at examinations 1 and 2 were 0.594 (standard error 0.060) for observer 1 and 0.581 (standard error 0.068) for observer 2, both of which correlate to moderate agreement.

Conclusion/Discussion

We found that it was possible to achieve good intraobserver reliability when experienced veterinarians assessed laryngeal function using the seven-grade system. Our findings compare favourably with intraobserver reliability when using a four-grade system, where mean κ statistics within each grade ranged from 0.65 (for grade 2) to 0.98 (for grade 4) and intraobserver agreement was 83.3% (Hackett *et al.* 1991). Although our percentage agreement is lower, with a four-grade system there is an increased likelihood that observers will agree by chance when compared to a seven-grade system. While κ statistics take agreement accounted for by chance into account, percentage agreements do not, so such percentages must be interpreted with care when comparing systems with different numbers of grades. We also found that it was possible to achieve good interobserver reliability using the seven-grade system. Our findings compare favourably with interobserver reliability when using a four-grade system, where κ statistics for concordance between 3 observers within each grade ranged from 0.51 (for grade 2) to 0.90 (for grade 4) (Hackett *et al.* 1991). In our study, both observers assigned the same grade in 63.1% of horses, which is similar to findings with four-level grading, where all observers assigned the same grade in 64% of horses (Hackett *et al.* 1991).

Intraobserver agreement was lowest when deciding betweeen grades 2.1, 2.2 and 3.1 and interobserver agreement was lowest when assigning a grade 2.1. This suggests the most difficult decision faced by observers was whether full arytenoid cartilage abduction was maintained once it was achieved. Hackett *et al.* (1991) found the majority of disagreements over grading were due to fluctuations between a grade 1 and 2 and therefore concluded that most disagreement existed over the degree of synchrony rather than the ability to reach full abduction. Deciding whether full arytenoid cartilage abduction is maintained once it has been achieved is a purely subjective assessment and may be a limitation of the seven-grade system.

At yearling sales and some pre-purchase examinations, endoscopic examinations are performed to confirm 'respiratory soundness' but it is uncertain how good a predictor this is of future respiratory function. Baker (1982) proposed that in the vast majority of horses, endoscopic examination of laryngeal function at rest remains unchanged throughout life. However, Dixon *et al*. (2002) demonstrated that 15% of a study population of 351 horses severely affected with RLN had shown evidence of progression in the degree of laryngeal dysfunction over a median of 12 months. In contrast, Anderson *et al.* (1997) reported that laryngeal function grade improved in 29% of horses and worsened in 28% of horses when endoscopy was performed at a 16 month interval. Proposed explanations for horses interchanging between what is considered 'normal' and 'abnormal' are poor repeatability of endoscopic grading and degeneration and regeneration of nerve fibres over. However an alternative explanation is that the endoscopic appearance of laryngeal function in any particular horse may not be consistent on a day-to-day basis. We found that when horses received 2 endoscopic examinations within 24-48 hours, only 41.7% of horses were assigned the same grade at both examinations, and the κ statistic indicated only moderate agreement between grades assigned to each horse at examinations 1 and 2. Changes in grades occurred in both directions with similar frequency, suggesting these changes were not simply due to the horses having adapted to the endoscopic procedure. Endoscopic examinations were performed through the right nostril since this made the examination easier to perform. To our knowledge there is only one previous study which has investigated intrahorse variation over a short time period, in which resting endoscopic examination was repeated on a group of 20 horses within 1 week (Ducharme *et al.* 1991). In this study a four-grade system was used, and the median laryngeal grade was found to be different in 21% of horses when endoscopy was performed via the right nostril on both occasions (Ducharme *et al.* 1991). It is unsurprising that a lower percentage of horses change in grade when a four-grade system is compared to a seven-grade system, since each grade is considerably broader and grades are more likely to agree by chance. Possible causes of arytenoid cartilage movement changing in an individual horse when endoscopy is performed on different days include a true variability in activity of the intrinsic laryngeal musculature, the influence of duration of examination and the influence of level of excitement during the examination. We suggest that single resting endoscopic examinations performed at time of sale should be interpreted with some caution. The importance of intrahorse variability can not be understated, as it has significant implications both clinically and for research into RLN.

In conclusion, the results of this study show that the seven-grade system proposed at the Havemeyer foundation has good intraobserver and interobserver reliability. We believe the clearly defined grading criteria provided by this system will help reduce variability between researchers and clinicians investigating RLN. The high amount of variability found to exist between grades assigned to a population of horses examined twice in a 24-48 hour period suggest that laryngeal function should not be assessed solely on a single examination, and that other factors, namely the presence of an inspiratory noise and palpable evidence of cricoarytenoid dorsalis muscle atrophy, must be considered concurrently.

References

Anderson, B.H., Kannegieter, N.J. and Goulden, B.E. (1997) Endoscopic observations on laryngeal symmetry and movements in young racing horses. *N. Z. vet. J.***45**, 188-192.

Baker, G.J. (1982) Laryngeal asynchrony in the horse: Definition and significance. In: *Proceedings of the 1st International Conference on Equine Exercise Physiology*, Eds. D.H. Snow, S.G.B. Persson and R.J. Rose, Granta Publications, Cambridge. pp 46-50.

Dixon, P.M., McGorum, B.C., Railton, D.I., Hawe, C., Tremaine, W.H., Pickles, K. and McCann, J. (2002) Clinical and endoscopic evidence of progression in 152 cases of equine recurrent laryngeal neuropathy (RLN). *Equine vet. J.***34**, 29-34.

Dixon, P.M., Robinson E., Wade, J.F. (2003) Havemeyer Foundation - Workshop Summary. In: *Proceedings of a Workshop on Equine Recurrent Laryngeal Neuropathy*, R & W Publications (Newmarket) Limited, Stratford-upon-Avon, UK. pp 93-97.

Ducharme, N.G., Hackett, R.P., Fubini, S.L. and Erb, H.N. (1991) The reliability of endoscopic examination in assessment of arytenoid cartilage movement in horses. II. Influence of side of examination, reexamination, and sedation. *Vet. Surg.***20**, 180-184.

Hackett, R.P., Ducharme, N.G., Fubini, S.L. and Erb, H.N. (1991) The reliability of endoscopic examination in assessment of arytenoid cartilage movement in horses. I. Subjective and objective laryngeal evaluation. *Vet. Surg.***20**, 174-179.

Grade	Description	Sub-grade
I	All arytenoid cartilage movements are synchronous and symmetrical and full arytenoid cartilage abduction can be achieved and maintained.	
II	Arytenoid cartilage movements are asynchronous and/or larynx is asymmetric at times but full arytenoid cartilage abduction **can** be achieved and maintained.	.1 Transient asynchrony, flutter or delayed movements are seen.
		.2 There is asymmetry of the rima glottidis much of the time due to reduced mobility of the affected arytenoid and vocal fold but there are occasions, typically after swallowing or nasal occlusion when full symmetrical abduction is achieved and maintained.
III	Arytenoid cartilage movements are asynchronous and/or asymmetric. Full arytenoid cartilage abduction **cannot** be achieved and maintained.	.1 There is asymmetry of the rima glottidis much of the time due to reduced mobility of the arytenoid and vocal fold but there are occasions, typically after swallowing or nasal occlusion when full symmetrical abduction is achieved but not maintained.
		.2 Obvious arytenoid abductor deficit and arytenoid asymmetry. Full abduction is never achieved.
		.3 Marked but not total arytenoid abductor deficit and asymmetry with little arytenoid movement. Full abduction is never achieved.
IV	Complete immobility of the arytenoid cartilage and vocal fold.	

Table 1: Grading system of laryngeal function performed in the standing unsedated horse, outlined at the Havemeyer foundation (Dixon *et al.* 2003)

European College of Veterinary Surgeons

Check session: (only one)

Scientific session – oral
- SA General
- SA Orthopedics
- LA General
- LA Orthopedic

Resident's Forum
- Large animal
- Small animal

Poster
- large animal
- small animal

Poster format A0
(width 90 / heighth 120 cm)

NOT for poster

Indicate in case you accept also to present your abstract as poster