Endoscopic scoring of the tracheal septum in horses - its reliability as a diagnostic tool and clinical relevance for the evaluation of lower airway health in horses

Inaugural-Dissertation

zur Erlangung der Doktorwürde
der Vetsuisse-Fakultät der Universität Bern

vorgelegt von

Christoph Koch
von Konstanz, D

2005
Von der Vetsuisse-Fakultät der Universität Bern auf Antrag von Prof. Dr. R. Straub als Dissertation genehmigt.

Bern, Der Dekan der Vetsuisse-Fakultät der Universität Bern
Vetsuisse-Fakultät der Universität Bern (2005)

Endoscopic scoring of the tracheal septum in horses - its reliability as a diagnostic tool and clinical relevance for the evaluation of lower airway health in horses.

Christoph Koch

Reasons for performing study: Although endoscopic scoring of the tracheal septum thickness is used as a diagnostic tool for evaluation of lower airway disease, its reliability and clinical relevance has never been critically assessed in the horse. Objectives: To investigate if septum thickness scores (STS) are a reliable and clinically useful measure in horses with lower airway disease. Methods: The variance of STS attributable to differences within and between observers, differences within individual horses as well as the variance attributable to changes over time was determined. Furthermore, the distribution of STS in a population of asymptomatic horses was investigated. Correlations of STS with age, gender, as well as mucus accumulation and cell differentials of tracheo-bronchial secretions and bronchoalveolar lavage fluid were evaluated. Effects on septum thickness as a consequence of altered pulmonary ventilation induced by different drugs (bronchospasm induced by metacholine, forced in- and exhalation provoked by lobelinhydrochloride and increased depth of breathing by romifidin hydrochloride) were assessed. Finally, STS of RAO patients were compared to those of clinically healthy horses. Results: STS showed an excellent intraobserver agreement, when performed by a “trained” observer, and satisfactory interobserver agreement. Established clinical, endoscopic and cytological measures of lower airway inflammation, i.e. mucus accumulation scores and airway neutrophilia, however, did not correlate with STS. In horses ten years and older septum scores were significantly higher than in horses younger than ten years of age. STS did not differ significantly between asymptomatic horses and RAO-affected horses both in exacerbation and in remission. Sedation with romifidine hydrochloride had no effect on STS. In horses with markedly increased breathing effort (e.g. obvious heaves, or after metacholine- or lobelinhydrochloride-challenge), endinspiratory and endexpiratory STS often differed dramatically, but the means of endinspiratory and endexpiratory scores were not different from baseline STS. Conclusions and clinical relevance: Endoscopic STS are a reproducible measure. However, even though previously used as an indicator of lower airway diseases, STS were not associated with clinical, endoscopic and cytological findings indicative of RAO or IAD. Therefore, there is no diagnostic significance of higher STS in patients suffering of lower airway diseases.

Keywords: horse; tracheal septum; endoscopic scoring; IAD; RAO
TABLE OF CONTENTS

1. LITERATURE REVIEW
   1.1 Description and definitions of equine lower airway diseases
   1.2 Endoscopy of the equine lower airway tract
   1.3 Endoscopic scoring systems
   1.4 Endoscopic scoring systems as a diagnostic tool in lower airway diseases

2. INTRODUCTION, AIMS AND QUESTIONS

3. MATERIALS AND METHODS

4. RESULTS

5. DISCUSSION AND CONCLUSIONS

6. MANUFACTURERS’ ADDRESSES

7. ACKNOWLEDGEMENTS

8. REFERENCES

9. SUMMARY

10. ZUSAMMENFASSUNG

11. FIGURES
1. LITERATURE REVIEW

1.1 Description and definitions of equine lower airway diseases

Various terms referring to non-infectious, environmentally induced equine lower airway disease exist. We will subsequently use the terms recurrent airway obstruction (RAO) and inflammatory airway disease (IAD), as defined at two recent, international workshops on inflammatory airway disease (Anon 2001, Anon 2002).

Clinical signs of horses affected with RAO are obvious and often dramatic when stabled and fed with hay. Today, the significance of RAO as a debilitating disease is undisputed and its’ pathogenesis is fairly well understood, showing striking similarities to the pathogenesis of human asthma. The typical clinical signs of RAO such as respiratory distress manifested by nostril flare and an increased abdominal lift, coughing and nasal discharge are mainly due to cholinergic bronchospasm accompanied by mucus and neutrophil accumulation (>20% broncho-alveolar lavage fluid [BALF] neutrophils) in the airways (Derksen et al. 1985; Robinson et al. 1993; Robinson et al. 1999; Robinson 2001; Gerber et al. 2004b). These symptoms become evident when horses are stabled and exposed to hay dust and other indoor-environment-irritants but clinical remission is often achieved within days to weeks when horses are kept at pasture and/or given corticosteroids (Gerber 1970; Gerber 1973; Murphy et al. 1980; Soma et al. 1987; Pearson and Riebold 1989; Jackson et al. 2000; Robinson et al. 2000).

In contrast to RAO, horses affected with IAD often show only mild or no clinical symptoms at all and, therefore, many times go undetected unless one specifically looks for the signs. In the past, the term IAD was used exclusively to refer to mild airway inflammation as seen in young racehorses.

We use the term IAD to describe horses of all uses and age-groups that show less severe, or no clinical symptoms of lower airway disease at all: horses, for instance, that may cough,
show abnormal nasal discharge or poor performance, but also horses without these clinical signs showing increased airway inflammation (>5% BALF neutrophils; > 20% Tracheo-bronchial secretions [TBS] neutrophils), mucus accumulation and/or airway hyperreactivity in the absence of clinical signs of respiratory distress (Anon, 2002; Gerber et al. 2005; Widmer 2005). Although IAD is a very common disease among stabled mature horses, the pathogenesis of IAD is less well understood than that of RAO (Sasse et al. 1985; Bracher et al. 1991; Gerber et al., 2003).

1.2 Endoscopy of the equine lower airway tract

Besides the clinical findings and historical data on stabling, feeding, coughing or poor performance, the endoscopic observations and accompanying cytological findings of TBS and BALF represent the basis for diagnostic evaluation of lower airway diseases in horses. At many equine clinics, the endoscopic examination of the trachea is therefore part of the routine examination of horses presented with respiratory problems. The few, but often revealing observations to be made on an endoscopic examination of the trachea include mucus amount and appearance as well as colour and vascularity of the tracheal mucosa. Furthermore, many experienced clinicians mention the association of (chronic) lower airway disease in horses and the thickening and blunting of the tracheal septum. Participants of the Havemayer Workshop in Boston (2002), for example, stated in a general consensus that: “When performing an endoscopic examination of the lower airways, the gross observations of mucus appearance and the thickening and blunting of the bronchial bifurcations should be noted, (…)”. A comprehensive literature review in German, French and English has been performed, but no information on what veterinary clinicians, pathologists or anatomists assume to be the cause of this observation, nor any substantial data confirming the consistency or clinical relevance of such findings, can be found.
1.3 Endoscopic scoring systems

The best established endoscopic scoring system in equine medicine is that of evaluating gastric lesions in horses with EGUS, as proposed by Murray et al. Other, endoscopic scoring systems to examine the upper respiratory tract (PLH (Burell 1985), DDSP (Dorsal Displacement of the soft palate) and ILH (idiopathic laryngeal hemiplegia) (Baker 1982)) are also fairly well established. In human medicine a multitude of visual scoring systems in all disciplines of “diagnostic imaging” have become important tools in practice as well as research. To only mention a few: the Lund-Mackay scores in CT, the Lindblad/Herfors system (synovial inflammation scoring) and Ayral system (cartilage lesion scoring) in arthroscopy, the Hetzel-Dent and the Los Angeles scale in gastrointestinal endoscopy or the bronchitis index in bronchoscopy.

In human medicine, numerous evaluation studies with the goal to compare reliability and clinical relevance of different endoscopic scoring systems were published in recent years, providing interesting insights on endoscopic scoring systems in general:

One must always consider that both the observer as well as the subject itself may be important sources of variance in any scoring system. Comparing inter- and intra-observer variability, the level of agreement usually showed to be as follows: intra-observer > inter-observer > inter-centre (Winkfield et al, 2003). Moreover, when comparing so called “expert” and “trainee” endoscopists, the group of experts showed greater consistency in scoring and better reproducibility than the group of trainees (Pandolfino, 2002). In terms of the subject, it is usually not possible to obtain corresponding in vivo measures (e.g., the actual measured thickness and blunting of the tracheal septum) to the scored variable (e.g., the tracheal septum thickness, as it appears on the screen). It is this lack of validation of endoscopic scores, which will always make the communication between clinicians and the interpretation and comparison of results between studies difficult and uncertain. It is also worth mentioning, that in a review of quantitative arthroscopy, composite scoring systems incorporating multiple
variables (e.g. lesion size and lesion site) seemed to be conceptually advantageous and showed the best content validity because they capture more aspects of the evaluated disease (Oakley, 2003).

1.4 Endoscopic scoring systems as a diagnostic tool in lower airway diseases

The bronchitis index (BI), as earlier mentioned, represents an endoscopic scoring system developed mainly as a research tool for rating lower airway inflammation in humans. The BI, as introduced by Thompson et al. in 1993, is a composite, semiquantitative scale for the assessment of airway inflammation and scores the visual appearance of erythema, oedema, secretions and friability from 0 = normal, to 3 = remarkably abnormal (Thompson et al., 1993). The blunting of bronchial bifurcations is integrated in the BI as a mild sign for oedema (score 1 for the parameter oedema). Although oedema and secretions consistently contributed less to the total BI when compared with the contribution of erythema (48 - 70%), they (unlike friability) proved to be a significant component of the visual appearance of lower airway inflammation.

In recent years considerable efforts were successfully made to standardize and assess the clinical relevance of evaluating mucus accumulation in the equine trachea (Dieckman 1987; Dixon et al, 1995; Gerber et al, 2003, 2004 a and b) as well as mucus appearance (Dieckmann 1987; Herholz et al, 2002; Gerber et al, 2004 a). These efforts contributed to the fact that, today, gross endoscopic observations of mucus amount and appearance in the trachea have become an undisputed diagnostic tool in evaluating the lower airways of mature horses.
2. INTRODUCTION, AIMS AND QUESTIONS

In addition to mucus accumulation and the appearance of the mucosa, the thickening of the tracheal septum is one of few characteristics the veterinarian can directly assess when performing an endoscopy of the lower airways. This provides important information for the examiner, complementary to the findings of the clinical exam and not dependent on further laboratory work-up. As discussed above, some endoscopic assessments may be limited regarding their clinical relevance and also reproducibility. When evaluating equine lower airways, it has recently been shown that only the endoscopic scoring of mucus accumulation, but not that of mucus colour or apparent viscosity are valid for endoscopic scoring, due to important observer variance in the latter two characteristics (Gerber et al, 2004 a).

The goal of this study was to investigate the reproducibility of septum thickness scores (STS) and its clinical relevance for evaluation of lower airway disease in horses. Therefore we aimed to answer the following questions:

1. Does intra- or interobserver variance significantly limit the reliability of STS?

2. Without changes in environment, what is the normal distribution of STS in a population of asymptomatic (healthy and/or subclinically IAD-affected) horses and how consistent are STS within asymptomatic individuals over a three week period of time?

3. Do STS correlate with TBS cytology, mucus accumulation scores, age or gender in a population of asymptomatic (healthy and/or subclinically IAD-affected) horses?

4. How are STS influenced by altered respiratory work and ventilation volumes when pharmacologically inducing forced in-/exhalation (with lobelinhydrochloride), bronchospasm (with metacholine) or light respiratory depression (with romifidine hydrochloride)?

5. Do STS differentiate between asymptomatic and RAO-affected horses?
3. MATERIALS AND METHODS

Definitions:
The term “tracheal septum” describes the anatomical structure that divides the trachea into the main bronchi at what is called the “tracheal bifurcation” or “carina tracheae”. In accordance with consensus definitions of a previous workshop (Anon, 2001), horses with RAO included in this study had to consistently develop respiratory distress (increased breathing effort and rate) when stabled and exposed to hay. The term IAD, as used in this study, does not refer to the condition in young racehorses but to older “chronic coughers”, as proposed as an expanded definition at the International Workshop on IAD (Anon, 2002). When stabled and exposed to hay, clinically IAD-affected horses coughed, but would not show signs of respiratory distress. The group of asymptomatic horses used in this study consisted of horses without any clinicopathological findings indicative of lower airway inflammation as well as subclinically IAD-affected horses with mild airway neutrophilia (≥20% neutrophils in TBS or ≥5% neutrophils in BALF). All horses taking part in this study were full sized riding horses (weighing 400 kg and more) of various breeds.

Overview of experimental protocols
We used data from protocol 1 to determine interobserver variance and gather first information on STS in defined RAO-patients in remission, as well as defined IAD-patients. To further assess the intraobserver correlation and gather information about the distribution of STS in a population of defined healthy and subclinically IAD-affected horses as well as changes in STS-distribution over three weeks, we used protocol 2.
To find out how STS are influenced by drug-induced bronchospasm as well as altered ventilation due to pharmacological effects of lobelinhydrochlorid or under sedation with romifidine hydrochloride we used protocol 3.
In order to investigate how STS were changing in defined RAO-horses in remission compared to those in the same horses in exacerbation we used protocol 4.

*Protocol 1: observer and horse variation in association with airway inflammation*

For this protocol, we resorted to unpublished data and video-material from a previous study (Gerber *et al.*, 2004 a). In the crucial protocol of this previous study (also protocol 1) the same type of instruments/materials and the same criteria to classify the health status of horses were employed, as we are using in all following protocols.

Tracheoscopic examinations (Olympus\(^1\) CF-0140L) were performed in 9 horses (2 without clinical signs, 4 with IAD and 3 with RAO in remission; ages ranging from 9 to 24 years) and repeated three times after six, twelve and 24 hours. The horses were stabled and not exercised at least one day prior to the first endoscopy. When stabled, they were bedded on shavings and fed concentrate and/or good-quality hay. Each endoscopic examination was recorded on videotape (Sony\(^2\) DVCAM DRS- 20MDP) and subsequently digitized (DVgate Motion [acquired in .avi]- version 2.2.00 DVgate Assemble [convert to .mpeg] - version 2.2.00 [copyright Sony\(^2\) Corp. 2000]).

TBS were collected at each time (0, 6, 12 and 24 h) by introducing a Teflon-coated PVC catheter through the working channel of the endoscope. Ten ml of phosphate-buffered saline (PBS) were instilled through the catheter into the lower part of the trachea and the PBS mixed with secretions was immediately recovered. From the recovered fluid, 100 and 200 \(\mu\)L aliquots were centrifuged for cytospin preparations. After locally anesthetizing the region of the carina and larger bronchi by instilling 50 ml of lidocain (2%) solution, the endoscope was wedged in the smallest accessible peripheral bronchus (Derksen *et al.* 1985) and during the final exam (24 h) BALF was collected. In doing so, six 50-ml aliquots of PBS were infused through the working channel of the endoscope and immediately recovered by suction. The recovered lavage fluid was pooled and 100 as well as 200 \(\mu\)L aliquots were centrifuged for cytospin preparations.
Using a hemacytometer, total cell counts in BALF were made. The TBS and BALF cytospin preparations were stained (with May-Grünwald-Giemsa) and differential cell counts were obtained. The differential cell counts were expressed as percent of total cells by counting 200 cells using standard morphologic criteria under a light-microscope.

The digitized video clips of the endoscopic examinations were viewed on personal computer screens and scored for accumulation, appearance and localization of mucus, as well as the blunting and thickness of the tracheal septum according to the continuous scales scoring system illustrated in Fig 1. Scoring was performed in randomized order independently by three blinded observers (VG = observer 1 = trainee, NR = observer 2 = trainee and CK = observer 3 = expert), and repeated by each observer at least three weeks after the first scoring series.

**Protocol 2: observer and horse variation in a population of clinically healthy horses**

A group of 98 horses (ages ranging from 3 to 18 years) was examined at two top-level training stables, one specializing in dressage horses and the other in show-jumpers. Each horse was examined (clinical exam, endoscopy and TBS cytology) on two different occasions (exams 1 and 2), in a three week interval. Fourteen of the 98 horses were excluded from the study, because of mild clinical symptoms (or missing data, 1 horse) and could therefore not be classified as healthy or subclinically IAD-affected horses. Consequently, our control group of healthy and subclinically IAD-affected horses consisted of 84 horses.

A detailed history of all respiratory (character, severity and onset of clinical signs, duration of response to environmental change) or potentially confounding problems was documented. Each horse was examined at rest, prior to the endoscopic examination and the type of breathing (nostril flare, abdominal lift; scoring according to (Robinson et al. 2000)), the respiratory and heart rate (breaths and beats per minute, respectively), any evidence of cough or nasal discharge and the rectal temperature were evaluated and documented. Mucosal colour
and capillary refill time (CRT) were also examined. The submandibular lymph nodes, the sensitivity of the pharynx/larynx area and the coughing reflex were assessed. Tracheal and chest auscultation were performed before and during a rebreathing exam (with a plastic bag of 20 l volume) and abnormal respiratory sounds were scored and recorded.

As in all 4 protocols, the videoendoscopic recording for each horse consisted of the tracheal descend in order to document the mucus accumulation and was stopped after placing the instrument at approximately 5 to 10 cm craniod to the tracheal carina, to depict the septum and both parting main bronchi in full view. Endoscopic scoring was later performed without knowledge of diagnosis from videotapes. In addition to mucus accumulation, septum thickness was scored according to the continuous scales scoring system illustrated in Fig 1. Scoring was performed by the same blinded observer (observer 3 = CK) and repeated once at least three weeks after the first scoring series. Mucus accumulation, mucus appearance as well as mucus localization scores were all made according to an analogue 5-grade scoring system we have previously validated (Gerber et al. 2004 a).

TBS was collected through a catheter inserted in the working channel of the endoscope. Ten ml of sterile saline were infused into the lumen of the trachea and recovered by suction. Of the recovered fluid 25, 50, 75 and 100 μl aliquots were centrifuged for cytospin preparations. After staining the cytospin preparations with May-Grünwald-Giemsa, differential cell counts were obtained from TBS by counting 200 cells using standard morphologic criteria under a light-microscope. Differentials were expressed as percent of total cells. Cell scoring was performed by the same person without knowledge of the identity of the animals and the protocol periods.
Protocol 3: STS in asymptomatic and RAO-affected horses in remission and in exacerbation, unsedated-sedated, after metacholine-stimulation (bronchospasm) and also lobelin-hydrochloride-stimulation (forced in-/exhalation).

To investigate the effects of altered ventilation as provoked by different drugs, we repeatedly performed tracheoscopy on ten horses. All ten horses (8 without clinical signs and 2 with RAO; ages ranging from 5 to 19 years) were stabled, fed concentrate and good-quality hay and/or haylage.

First of all, each horse was examined before and after sedation. For sedation, 0.04 mg/kg romifidine hydrochloride (Boehringer Ingelheim³, Sedivet®) were applied intravenously. The septum thickness was then scored from videotape according to the continuous scale shown in Fig 1. All scoring was performed by the same observer (observer 3 = CK), not knowing the identity of the horse or whether the animal was sedated or not. STS was repeated once at least three weeks after the first scoring series.

Secondly, each of the ten horses was stimulated with inhaled nebulized metacholine chloride (Christoffel-Apotheke⁴). After sedating each horse with 0.04 mg/kg romifidine hydrochloride applied intravenously, an equine inhalation mask (EMMS⁵ Equine Face Mask – size 03, Part No EFM 103, as provided with the EMMS Equine oscillation system) was put on, tightly sealed and connected to a double-valve inhalation system with a compressor nebulizer for medical aerosols (AS1 PRO by Medel⁷). Metacholine-challenge sequences were started by nebulizing 1ml of sterile saline solution and followed by metacholine chloride diluted in saline solutions (1ml) at increasing concentrations (0.15 mg/ml, 0.5 mg/ml, 1.5 mg/ml, and 4.5 mg/ml and 13.5 mg/ml). Aerosolized saline solutions were administered over 2 minutes at each time and time intervals between each nebulization were 2 minutes or more. The metacholine nebulization was stopped when the horse showed signs of acute respiratory distress or when the highest dose of metacholine was reached. We then immediately removed the mask and repeated tracheoscopy while the horse was still showing a markedly increased
abdominal breathing effort. In seven (5 without clinical signs and 2 with RAO in remission; ages ranging from 5 – 19 years) of the ten horses we were successful in taking off the mask and performing tracheoscopy fast enough to capture the tracheal septum on videotape while the horse was still breathing hard. Because of the clearly remarkable differences in septum thickness during increased breathing, two STS were given at each tracheoscopy (one endinspiratory and one endexpiratory).

On another occasion, six (5 without clinical signs and 1 with RAO; ages ranging from 5 – 19 years) of these ten horses were given 0.2 mg/kg lobelinhydrochloride (G. Streuli & Co. AG Pharmaceutica, Lobelin Hydrochloridum) intravenously and the tracheal septum of each horse was recorded with a videoendoscope immediately before and during lobelinhydrochloride-stimulation. When stimulated with lobelinhydrochloride, horses again showed obvious differences in septum thickness during in- and exhalation. Tracheal STS were therefore paired, one end-inspiratory and one end-expiratory score, as described above for metacholine-stimulation.

**Protocol 4: horse variation in a population of defined RAO-horses in remission and in exacerbation**

In seven horses (ages ranging from 8 to 19 years), all showing clinical signs of RAO (i.e. respiratory distress when stabled and fed with hay and remission when removed from hay and stable environment), tracheoscopic examinations were performed when in remission and during exacerbation. A horse was defined as suffering from “RAO in exacerbation” when showing unmistakable symptoms of respiratory distress and airway obstruction during a period of environmental challenge by stabling in stalls with straw beddings and feeding hay. All horses were defined as having “RAO in remission”, after being stabled in a barn complex especially adapted to the requirements of diagnosed RAO patients (TierWohl beddings, haylage and concentrate feeding exclusively), for a period of at least three days or as long as it
would take until clinical symptoms of respiratory distress disappeared. As specified in *protocol 2*, a detailed history of all respiratory (character, severity and onset of clinical signs, duration of response to environmental change) or potentially confounding problems was documented and each horse was examined at rest, prior to the endoscopic examination. In addition to the clinical examination as described in *protocol 2*, we also determined the arterial partial oxygen pressure prior to every endoscopic examination. TBS and/or BALF were collected and evaluated for each endoscopy, as described in *protocol 1*.

All tracheoscopic examinations were documented on digital videotape and scores for mucus-accumulation as well as septum thickness were later given for each examination. The scoring was performed by the same observer (observer 3 = CK), unaware of the horses’ identity and health status. STS was repeated once at least three weeks after the first scoring series took place.

*Statistics*

NCSS\textsuperscript{9} 6.0.22, statistical analysis and data analysis software was used for all statistical analysis.

After accepting normal distribution of scored data (Martinez-Iglewicz test for normality), data from *protocol 1* were analyzed by means of a four-way ANOVA according to the model:

\[ Y = \mu + H + O + T + R + HO + HT + OT \]

where \( Y \) was the response variable (i.e. STS), \( \mu \) was the overall mean, \( H \) was the random effect of horse, \( O \) was the random effect of observer (inter-observer), \( T \) was the random effect of time, and \( R \) was the error due to repetition (intra-observer) and \( HO, HT \) and \( OT \) were interactions of those random factors. When solved for \( Y \), interactions of random factors \( HT \) and \( OT \) did not contribute to the variation. Variance \( (s^2 = \text{standard deviation}^2 = \sigma^2) \) was calculated by setting the Estimated Mean Square equal to the Mean Square calculated by
ANOVA, and solving for $\sigma^2$. Because there was no significant effect of time when it was considered a fixed factor, it was considered a random factor for the purpose of calculating $\sigma^2$.

To analyze the distribution of septum scores in a population of clinically healthy horses (protocol 2), we used NCSS$^{10}$ descriptive statistics and normal distribution was accepted for all performed normality tests. The number of observations (counts), the median as well as its’ interpercentile range were reported. Correlation of septum scores over three weeks within individuals, before and during lobelinhydrochloride-stimulation, before and after metacholine-stimulation, with and without sedation as well as the correlation of septum scores and mucus amount or TBS and BALF cell differentials were calculated by means of the Spearman Rank Order Test and results are shown as a correlation coefficient “r”.

We used Wilcoxon Rank-Sum Tests to compare unpaired, nonparametric data such as septum scores in different age-categories, septum scores in males and females and septum scores in defined clinically healthy or subclinically IAD-affected horses compared to those in age-matched RAO horses in remission or exacerbation respectively. Nonparametric, paired observations, like septum scores in RAO horses in remission vs. septum scores of the same horses in exacerbation or mean inspiratory vs. expiratory septum scores under metacholine- and lobelinhydrochloride-stimulation were tested, using Wilcoxon Signed Rank Tests.

Paired, continuous variables like partial arterial oxygen pressure, neutrophil-percentages in TBS or BALF, were assessed with Student’s paired T-Tests. Significance limit for all tests was set at $P < 0.05$. 

4. RESULTS

Observer, subject and time all contributed to the variance in STS. The subject or between horse variance ($\sigma^2_H$) contributed the most with 57% of the total variance. The variance attributable to repetitions (intra-observer variance; $\sigma^2_R$) and attributable to differences between observers (inter-observer variance; $\sigma^2_O$) proved to be almost the same, at which $\sigma^2_R$ was 20% and $\sigma^2_O$ just slightly higher with 22%. Time (i.e. horse over time variance; $\sigma^2_T$) had only very little influence on the variance of septum scores and only accounted for 1% of the total variance.

Mean septum scores (recorded twice over three weeks and each scored twice by the same observer $3 = CK$, see also materials and methods) in 84 horses were distributed normally by statistical criteria (Shapiro-Wilk W-, Kolmogorov-Smirnov -Test), with a median of 3.35 and an interquartile range of 0.8 (Fig. 2).

STS over a period of three weeks correlated ($r = 0.80$) well within individuals (Fig. 4). When categorizing horses into two age groups, horses age 10 years and older ($n = 15$ out of 84) showed higher STS ($P = 0.022$) compared with horses younger than ten years of age ($n = 69$ out of 84), as illustrated in Figs. 5 A and B. Gender had no significant influence on the STS ($P = 0.29$).

STS in individual horses with and without sedation correlated well ($r = 0.89$). The means of endinspiratory and endexpiratory STS during metacholine- or lobelinhydrochloride-stimulation correlated closely with the baseline STS given “at rest” ($r_{MCh} = 0.99$, $r_{lobel} = 0.94$). However, within individual horses there was a significant difference between endinspiratory and endexpiratory STS at forced in- and exhalation during metacholine- as well as lobelinhydrochloride-challenge ($P_{MCh} = 0.018$, $P_{lobel} = 0.035$). STS were up to 1.9 scoring points lower at forced inhalation than those at forced exhalation (metacholine-
challenge), and means of STS at forced inhalation versus exhalation differed 0.86 or 0.45 scoring points for metacholine- or lobelin-challenge, respectively (Figs. 6 A and B).

We found no correlations between STS and mucus accumulation scores ($r = 0.05$) or between TBS cell differentials and STS (STS and neutrophil-percentages: $r = 0.12$, STS and macrophage-percentages: $r = -0.17$, STS and lymphocyte-percentages: $r = -0.07$).

Septum scores in RAO horses both in remission and exacerbation showed no significant difference when compared to septum scores of age-matched (+/- 1 year) and otherwise randomly picked asymptomatic horses ($P_{\text{remission}} = 0.41$ and $P_{\text{exacerbation}} = 0.65$, respectively).

As illustrated in Fig. 8A, the thickening of the tracheal septum in exacerbation compared to remission was minimal (difference in means: 0.12 scoring points, and maximal difference in scoring points was 0.4) and showed only a trend for significance ($P = 0.16$ two-tailed; $P = 0.08$ one-tailed).
5. DISCUSSION AND CONCLUSIONS

Our perspective in evaluating the thickness of the tracheal septum was that of the clinician and our approach therefore focused on the correlation of the visual score with the clinical symptoms as well as established additional endoscopic and cytological measures. Since to our knowledge neither the relevance of a thickened or blunted tracheal septum, nor the reliability of STS have ever been investigated, we first of all had to validate our scoring system.

Are STS a reproducible measure?

The subject, as hypothesized, attributed most to the total variance of septum scores. Within-horse variance over time, when examined four times within 24 hours, contributed a mere 1% to the total variance. It is important to note, however, that our results also show the influence of the observer. Based on the differences we saw between observers, one must expect the mean scores to vary significantly, but not dramatically by approximately +/- 0.9 scores between observers (results not shown). Observer related variance, therefore, contributed a fairly high 42% (intra- and interobserver combined). Interpreting this strong observer contribution several circumstances must be taken into account: Observers 1 and 2 can be considered “trainee observers” compared to observer 3. This difference in training is well reflected in the obvious differences in correlation coefficients of the three observers scoring in protocol 1 (r_{observer 1} = 0.43, r_{observer 2} = 0.62, r_{observer 3} = 0.97; results not shown above).

Comparing the correlation coefficients of the three observers, it is also of note that observer 3, demanding higher quality of the video material scoring was performed with, did not score as many septums as the other observers. Thus, when observer 2 scored all but one septum in scoring-round 1 and all septums in scoring-round 2 (providing a total of 35 paired scores to be correlated), observer 3 provided only a total of 24 paired septum scores and considered the rest of the digitized video sequences as “insufficient” for scoring, because either the septum
couldn’t be viewed long enough or visibility was poor. When endoscopic scoring of the tracheal septum was assessed in protocol 2 (n = 84), the endoscopic examination was performed by observer 3 himself. Therefore, the digitized video material met all quality criteria: Most importantly, the septum had to be visible long enough and the endoscope had always to be placed at approximately the same distance to the carina, providing an image of the septum and both parting main bronchi as depicted in Figs. 1, 6 and 7. Consequently, all 84 paired scores as repeated by observer 3 after three weeks or more, revealed a very satisfactory correlation coefficient (r = 0.89, Fig. 3).

Considering the better quality of video material provided for protocol 2, the better correlation coefficient of observer 3 in protocol 1 compared to his correlation coefficient for protocol 2, seems contradictory. This, however, presumably reflects the wider variability of septum thickness (respectively the greater rage of septum scores) in protocol 2 as well as the unequal composition of subjects scored in protocols 1 and 2. The latter needs further explanation in extenso: The subjects scored in protocol 2 showed much greater heterogeneity than the subjects scored in protocol 1. As earlier described in detail, in protocol 1, only nine horses (n = 9) were examined, but each horse was examined repeatedly four times within 24 hours, providing 4 scores per scoring-round. In protocol 2, in contrast, a total of 84 different horses (n = 84) were examined and scored, each horse providing only one score per scoring round. Presumably due to an effect of recognition, scores assessed by an observer agree better when repeatedly scoring a more or less identical subject four times than when repeatedly scoring 4 different subjects.

Considering the strong contribution of the actual object of interest (i.e. in practice the differences of STS between horses) on the total variance and the excellent correlation of repeated scores for observer 3, we conclude that STS proved to be a reproducible measure, especially for trained observers.
STS in a population of asymptomatic horses

The distribution of mean septum scores in 84 clinically healthy Warmblood horses fulfilled all criteria of a Gaussian distribution (Fig. 2). Septum scores within individuals correlated very well ($r = 0.80$) when repeating endoscopic examination in asymptomatic horses after three weeks (Fig 4). This indirectly confirms that variance in repeatedly performed STS is influenced by changes in septum thickness over time, as shown in protocol 1 ($\sigma^2_T = 1\%$) not even over a time period of three weeks.

When assessing the influence of age, we categorized horses in groups younger and older than 10 years, because in our clinic more than 50% of all horses diagnosed with RAO in the past seven years were 10 years and older (Conrot and Gerber, unpublished data). Therefore, we particularly wanted to avoid misinterpreting thicker septum scores in those horses at higher risk of having RAO. Surprisingly enough, asymptomatic horses of age 10 years and older consistently had thicker septums than younger (<10) asymptomatic horses (Figs. 5 A and B). This finding could possibly explain why some clinicians have the impression, that RAO-affected horses have a tendency for thicker septums than asymptomatic horses or horses with IAD, since more severe forms of lower airway inflammation are more frequently seen in older horses.

To determine whether STS could be directly correlated with established measures of lower airway inflammation, we compared septum scores with neutrophil-percentages in TBS as well as endoscopic mucus accumulation scores. As shown in the results section, no correlation was found for any of these measures. Here, it must be pointed out that merely one third of all examined (stabled) horses, even though asymptomatic, could be classified as “healthy” regarding their lower airways, when applying the criterion of TBS neutrophil-percentages being 20% or less [as recently proposed on international workshops on inflammatory airway disease (Anon 2001, Anon 2002)]. Furthermore, it shows that STS cannot be used to distinguish different degrees of subclinical IAD. In addition, we also compared STS with
motivation- and performance-parameters of these horses as rated by riders and again found no correlation at all (results not shown).

*Septum scores and alterations of pulmonary ventilation*

When planning protocol 3 (septum scores before and during drug-induced dynamic deformation of the lower airways) we first of all wanted to determine if sedation alters STS, as it is necessary for some horses to be sedated when performing a thorough endoscopic examination. Our results clearly demonstrate, that this is definitely not the case, since septum scores before and during sedation correlate extremely well ($r = 0.89$). Second of all, we also hoped to gain information about what may cause the obvious differences in STS observed in different horses, regardless of their health-status.

Still, we can only speculate about what causes the noticeable differences in thickness of the tracheal septum in different horses and (less obvious) within individual horses. The hypothesis, however, that active bronchospasm exclusively is responsible for the visible changes in STS within individual horses as well as the (so far presumed) differences between healthy horses and horses with lower airway disease, must be rejected: Metacholine-response does not significantly alter the mean STS, but only leads to distinguishably different septum scores during in- and exhalation, due to an increased breathing effort provoked by lower airway obstruction. Bronchospasm, however, persists over in- as well as exhalation and should therefore have a significant influence on the arithmetical mean of in- and expiratory septum scores compared to the septum score without metacholine-stimulation. Furthermore, when inducing forced in- and exhalation with lobelinhydrochloride, which does not provoke bronchospasm but acts as a respiratory stimulant, the same phenomenon of obviously altering STS during in- and exhalation can be observed.

The observation that STS during a markedly increased breathing effort, as in severely heaves-affected horses or induced with metacholine or lobelinhydrochloride, showed significant
differences at in- and exhalation of up to 1.9 scoring points (Figs.6 and 7), as well as the finding that horses older than 10 years tend to have thicker tracheal septums, lets us suggest a different hypothesis: The clearly visible differences in septum thickness between different horses represent a physiological morphologic variation. This variation of septum thickness is presumably influenced by many different factors such as the individual horses’ constitution (radius and shape of the tracheal and bronchial cartilages, angle between the separating main bronchi, individual thorax-shape) as well as pulmonary compliance. To support this theory, we conducted a further experiment with an isolated lung of a euthanized horse (without signs of any lower airway diseases) and were again able to demonstrate that the septum appears much thicker when the lung is completely collapsed compared to the septum after the lung was inflated with compressed air (Fig. 7). To determine the true nature of this phenomenon, however, would demand further experimental and patho-morphological investigations that would go beyond the scope of this study.

Clinical relevance of STS

Although not statistically significant, there was a trend (P = 0.08, one-tailed hypothesis) for STS to increase when comparing scores in defined RAO horses in remission to RAO horses in exacerbation. It must be noted, however, that these changes in STS, apply exclusively to the particular individual scored in exacerbation and again in remission. As indirectly proposed by Thompson et al., it may be due to oedema that the tracheal septum sometimes appears slightly thickened within the same individuals once active inflammation has been provoked. These mild increases in STS, however, are so subtle (a maximal difference of 0.4 scoring points and mean difference of only 0.12 scoring points) that even a trained examiner would have difficulties in confidently distinguishing an increase in STS. The established clinicopathologic findings such as neutrophil-percentages in TBS or BALF and arterial partial oxygen pressure (Figs.8 A - D) are much more reliable and sensitive tools when analyzing the
activity of lower airway inflammation and its clinical and functional consequences for RAO patients.

Although oedema cannot explain the obviously visible differences in tracheal septum thickness observed between different horses (affected with lower airway diseases or not) the evaluation of oedema in lower airway inflammation may potentially still become valuable when creating a reliable and useful endoscopic grading system. Providing that oedema is better perceptible when endoscopically evaluating the blunting and thickening of smaller bronchial bifurcations (as proposed by Thompson et al., 1993), oedema, abnormal mucus production as well as other visible signs of inflammation (like erythema) could be integrated in a comprehensive composite endoscopic scoring system on lower airway inflammation, similar to the BI in human medicine.

Conclusions
Tracheal STS are a reproducible measure within the same and between different observers. However, we have also come to the conclusion that tracheal STS are irrelevant as a single measure when attempting to differentiate patients with chronic or acute inflammation and/or hyperresponsiveness of their lower airways from patients with healthy lower airways. In particular, tracheal STS did not show any correlation with clinical or clinicopathologic findings indicative of RAO or IAD. For these reasons we state that: Endoscopic scoring of the tracheal septum in horses, although shown to be a reproducible measure, is not a clinically useful diagnostic tool to help differentiate horses affected by lower airway diseases from horses with healthy lower airways.

However, dynamic alterations in STS indirectly reflect disturbances in respiratory work and functional impairment of pulmonary ventilation and could be used to document and quantify severe respiratory distress.
6. MANUFACTURERS’ ADDRESSES

1Olympus Optical (Schweiz), Chriesbaumstr. 6, CH-8903 Schwerzenbach, Switzerland.

2Sony Corp., Sony Europe Headquarters, Kemperplatz 1, Berlin, Germany.

3Boehringer Ingelheim (Schweiz) GmbH, CH-4002 Basel, Switzerland.

4Christoffel-Apotheke, Christoffelgasse 3, CH-3011 Bern, Switzerland.

5G. Streuli & Co. AG Pharmaceutica, Bahnhofstrasse 7, CH-8730 Uznach, Switzerland.

6EMMS, Unit 12 Woolmer Way, Bordon, Hants, United Kingdom.

7Medel S.p.A., Via Micheli 7, 43056 San Polo di Torrile, Italy.

8TierWohl, Rudolf Speiser, Aeschbrunnhof 75, CH-4469 Anwil, Switzerland.

9NCSS Statistical Software, 329 North 1000 East, Kaysville, Utah 84037, USA.

7. ACKNOWLEDGEMENTS

I especially want to thank my wife Sibyl and two children Amelie and Lotta for their never ending patience and support throughout this study. All co-workers Annina Widmer, Alessandra Ramseyer, Thea Rhyner, Evelyn Bussmann, Ursi Jost and my coaches V. Gerber and R. Straub, for all their help and commitment. I also want to express my special gratitude to all owners of RAO horses that participated in this study, as well as U. Kasselmann and P. Schockemöhle for the permission to examine their horses and their teams for their kind help.
8. REFERENCES


9. SUMMARY

*Reasons for performing study:* Although endoscopic scoring of the tracheal septum thickness is used as a diagnostic tool for evaluation of lower airway disease, its reliability and clinical relevance has never been critically assessed in the horse.

*Objectives:* To investigate if septum thickness scores (STS) are a reliable and clinically useful measure in horses with lower airway disease.

*Methods:* The variance of STS attributable to differences within and between observers, differences within individual horses as well as the variance attributable to changes over time was determined. Furthermore, the distribution of STS in a population of clinically healthy horses was investigated. Correlations of STS with age, gender, as well as mucus accumulation and cell differentials of tracheo-bronchial secretions and bronchoalveolar lavage fluid were evaluated. Effects on septum thickness as a consequence of altered pulmonary ventilation induced by different drugs (bronchospasm induced by metacholine, forced in- and exhalation provoked by lobelinhydrochloride and increased depth of breathing by romifidin hydrochloride) were assessed. Finally, STS of RAO patients were compared to those of clinically healthy horses.

*Results:* STS showed an excellent intraobserver agreement, when performed by a “trained” observer, and satisfactory interobserver agreement. Established clinical, endoscopic and cytological measures of lower airway inflammation, i.e. mucus accumulation scores and airway neutrophilia, however, did not correlate with STS. In horses ten years and older septum scores were significantly higher than in horses younger than ten years of age. STS did not differ significantly between asymptomatic horses and RAO-affected horses both in exacerbation and in remission. Sedation with romifidine hydrochloride had no effect on STS. In horses with markedly increased breathing effort (e.g. obvious heaves, or after metacholine- or lobelinhydrochloride-challenge), endinspiratory and endexpiratory STS often differed
dramatically, but the means of endinspiratory and endexpiratory scores were not different from baseline STS.

**Conclusions and clinical relevance:** Endoscopic STS are a reproducible measure. However, even though previously used as an indicator of lower airway diseases, STS were not associated with clinical, endoscopic and cytological findings indicative of RAO or IAD. Therefore, there is no diagnostic significance of higher STS in patients suffering of lower airway diseases. However, dynamic alterations in STS indirectly reflect disturbances in respiratory work and functional impairment of pulmonary ventilation and could be used to document and quantify severe respiratory distress.
10. ZUSAMMENFASSUNG

*Beweggrund für die Studie:* Obwohl die endoskopische Beurteilung des Trachealseptums bereits als diagnostisches Hilfsmittel seine Anwendung findet, wurde seine klinische Aussagekraft und Reproduzierbarkeit, nach unserem Kenntnisstand, noch nie systematisch untersucht.

*Fragestellung:* Genügt die endoskopische Beurteilung der Dicke des Trachealseptums in Form eines „Scores“ auf einer kontinuierlichen Skala von 1 bis 5, als zuverlässig reproduzierbarer und klinisch aussagekräftiger Parameter für die tracheoskopische Untersuchung von Pferden?


*Resultate:* Die Streuung von wiederholt erhobenen Septum-Scores setzt sich aus folgenden Faktoren zusammen: Eigentliche Unterschiede des zu beurteilenden Subjekts (56%), mangelnde Übereinstimmung von wiederholten Scores durch verschiedene Beobachter (22%) und ein und des selben Beobachters (20%), sowie Veränderungen des Subjekts im Verlauf der Zeit (1%). Ein geübter Beobachter, jedoch, zeigt eine sehr gute Übereinstimmung bei

11. FIGURES

Fig. 1:

Fig. 2:
Fig. 3:

![Scatter plot with an r value of 0.89 showing the correlation between STS - Round 1 and STS - Round 2.]

Fig. 4:

![Scatter plot with an r value of 0.80 showing the correlation between Mean STS 1st Trip and Mean STS 2nd Trip.]

Fig. 5:

A

Age in years

STS

1 2 3 4 5

3-9 years

10 years and older

B

P = 0.02

STS

3.3 3.4 3.5 3.6 3.7 3.8 3.9 4.0 4.1 4.2 4.3 4.4 4.5 4.6 4.7 4.8 4.9 5.0

3-9 years

10 years and older
**Fig. 6:**

A. STS (P = 0.02)

B. STS (P = 0.04)

Inhalation - Exhalation

**Fig. 7:**

A. Heaves-affected

B. Post mortem specimen

Inflated - Deflated

Inhalation - Exhalation
Fig. 8:

A

STS

remission – exacerb.

P = 0.16

B

Arterial pO2 (mm Hg)

remission – exacerb.

P = 0.003

C

TBS-neutrophil-percentages

remission – exacerb.

P = 0.034

D

BALF-neutrophil-percentages

remission – exacerb.

P = 0.015
Figure legends

Fig. 1: Endoscopic septum thickness scores

Fig. 2: Distribution of septum thickness scores (STS) in asymptomatic horses (n = 84), mean of both exams

Fig. 3: Intraobserver (observer 3) correlation of septum thickness scores (STS): round 1 represents the first scoring of endoscopic recordings and round 2 represents the repeated scoring three weeks after round 1. R is the Spearman Rank Order correlation coefficient.

Fig. 4: Correlation of septum thickness scores (STS) within individuals over three weeks. The mean septum scores of both scoring rounds for exam 1 are compared to the mean septum scores of both scoring rounds for exam 2. Exam 2 was performed three weeks after exam 1. R is the Spearman Rank Order correlation coefficient.

Fig. 5: Septum thickness scores (STS) and age in a population of 84 clinically healthy horses (A) and presented in a box plot showing horses younger than 10 with significantly (P=0.02) lower STS than horses of 10 years of age and older (B).

Fig. 6: Septum thickness scores (STS) and corresponding endoscopic views of the tracheal septum under metacholine(A) - and lobeline(B) - challenge: at forced inhalation (left) versus forced exhalation (right) for 7 and 6 horses, respectively.
Fig. 7: Endoscopic recording of the tracheal septum in a severely heaves-affected horse at forced inhalation (left) and forced exhalation (right) (A). Endoscopic recording of a tracheal septum in an isolated lung of a euthanized horse after (left) and before inflation (right) with compressed air (B).

Fig. 8: Endoscopic septum thickness scores (STS, A), arterial partial oxygen pressure (B), TBS- and BALF-neutrophil- percentages (C and D, respectively) of RAO horses in remission and in exacerbation. STS in remission and exacerbation did not differ significantly ($P=0.16$, not significant), whereas arterial partial oxygen pressure was significantly ($P=0.003$) higher in remission and TBS- as well as BALF-neutrophil-percentages were significantly ($P=0.034$ and $P=0.015$, respectively) lower in remission than in exacerbation.