Tracheobronchoscopic assessment of exercise-induced pulmonary hemorrhage in horses

Kenneth W. Hinchcliff, BVSc, PhD; Melissa A. Jackson, BSc; James A. Brown, BVSc; Anthony F. Dredge, BVSc; Paul A. O’Callaghan, BVSc; John P. McCaffrey, BVSc; Paul S. Morley, DVM, PhD; Ronald F. Slocombe, BVSc, PhD; Andrew F. Clarke, BVSc, PhD

Objective—To determine the interobserver variability of assessment of exercise-induced pulmonary hemorrhage (EIPH) during tracheobronchoscopic examination in horses.

Animals—747 Thoroughbred racehorses.

Procedures—850 tracheobronchoscopic examinations were performed within 2 hours of racing for the horses. Examinations were recorded on videotape, and EIPH and its severity were assessed independently by 3 veterinarians. Concordance was determined by calculation of the Cohen weighted \( \kappa \) statistic and tabulation of scores assigned by each observer.

Results—Weighted \( \kappa \) statistics ranged from 0.75 to 0.80. In 99.4% of observations, all observers agreed or 2 of 3 agreed and the third differed by ≤1 grade.

Conclusions and Clinical Relevance—Results indicated that interobserver reliability of tracheobronchoscopic assessment of EIPH in Thoroughbred racehorses is high when the examination is conducted by experienced veterinarians. Concordance among investigators is sufficient to justify use of this grading system for further studies and clinical descriptions of EIPH. (Am J Vet Res 2005;66:596–598)

Exercise-induced pulmonary hemorrhage (EIPH) is a pervasive condition of Thoroughbred racehorses. Determination of the importance of EIPH in terms of athletic performance, well-being of horses, investigation of the role of putative risk factors, and examination of the efficacy of treatment requires a method of assessment of severity of hemorrhage that is highly repeatable among observers. Techniques used to assess the severity of EIPH include measurement of RBC count in bronchoalveolar lavage fluid or subjective assessment of the severity of hemorrhage based on the presence of blood in airways detected during tracheobronchoscopic examination. Tracheobronchoscopic assessment of severity of EIPH offers the advantages of simplicity and immediacy in that handling and laboratory processing of samples are not necessary. However, previous descriptions of the use of tracheobronchoscopic assessment of severity of EIPH have not reported on the interobserver reliability of the technique. The purpose of the study reported here was to determine the interobserver reliability of tracheobronchoscopic assessment of severity of EIPH in Thoroughbred racehorses.

Materials and Methods

Study design—Eight hundred fifty tracheobronchoscopic examinations were performed on a convenience sample of 747 Thoroughbred racehorses within 2 hours of racing. Races were 1,000- to 3,200-m flat races run on turf at Flemington, Moonee Valley, Caulfield, and Sandown racetracks in Melbourne, Australia, from March 1 to June 18, 2003.

Tracheobronchoscopic examinations of the upper airways, pharynx, larynx, trachea, and carina were performed on unsedated horses restrained by use of a nose twitch and halter. The nasopharynx and trachea to the level of the tracheal bifurcation were examined for the presence of blood. The examination was recorded on videotape for subsequent review. The videotaped examinations were independently reviewed by 3 veterinarians (JAB, AFD, KWH), and the severity of EIPH was graded according to modifications of described grading systems that had 4 or 5 grading levels. For this study, grade 0 = no blood detected in the pharynx, larynx, trachea, or mainstem bronchi; grade 1 = presence of 1 or more flecks of blood or 2 or fewer short (less than one quarter the length of the trachea) streams of blood in the trachea or mainstem bronchi visible from the tracheal bifurcation; grade 2 = 1 long stream of blood (greater than half the length of the trachea) or more than 2 short streams of blood occupying less than a third of the tracheal circumference; grade 3 = multiple, distinct streams of blood covering more than a third of the tracheal circumference, with no blood pooling at the thoracic inlet; and grade 4 = multiple, coalescing streams of blood covering more than 90% of the tracheal surface, with blood pooling at the thoracic inlet (Figure 1).

Statistical analyses—Weighted \( \kappa \) statistics were calculated for each combination of observers. Weighted \( \kappa \) statistics take into account the magnitude

Unauthenticated | Downloaded 06/04/22 08:08 PM UTC
of the difference between observers for a particular observation. Disagreement between observers of 1 grade is considered of less importance than is disagreement by 2 or more grades. The \( \kappa \) statistics were considered to indicate poor (\( \kappa < 0.20 \)), fair (0.20 to 0.40), moderate (0.41 to 0.60), good (0.61 to 0.80), and excellent (> 0.80) agreement. Results are reported as mean and 95% confidence intervals (CIs).

**Results**

Prevalence of EIPH, defined as blood in the trachea or bronchi detected by at least 1 observer, was 68.4%. Weighted \( \kappa \) statistics for concordance between pairs of observers were 0.76 (95% CI, 0.73 to 0.79), 0.80 (95% CI, 0.77 to 0.83), and 0.75 (95% CI, 0.72 to 0.79). Severity scores were identical for all 3 observers for 68.7% of examinations. The 3 observers concurred on assignment of 269 examinations as grade 0, 227 as grade 1, 63 as grade 2, 17 as grade 3, and 8 as grade 4. Scores of 2 of the 3 observers agreed and that of the third observer differed by 1 grade in 30.7% of examinations. Scores of 2 of 3 observers agreed and that of the third observer differed by ≥ 2 grades in 0.4% of observations. All 3 observers disagreed in 0.2% of examinations. Therefore, all observers agreed or 2 of 3 agreed and the third differed by ≤ 1 grade in 99.4% of observations.

**Discussion**

Results of the study reported here indicate that it is possible to achieve a high degree of concordance among experienced investigators for assessment of
severity of EIPH in Thoroughbred racehorses. Previous investigators using similar techniques have not provided measures of the interobserver variability of their assessment of EIPH. Quantification of interobserver variability is important if grading systems such as those used here are to be used by more than 1 observer in either clinical or experimental situations. Our study has important implications for the study of EIPH in that it validates the use of a subjective scoring system, thereby providing a tool for investigation of the causes and effects of EIPH. A weakness of our study was that although it provides an ordinal assessment of the severity of EIPH, it does not provide a quantitative measure of the amount of hemorrhage.

The concordance between observers was good, as indicated by \( \kappa \) statistics > 0.75. However, use of the \( \kappa \) statistic to assess concordance among observers is problematic when the prevalence of the condition is high. This occurs because the \( \kappa \) statistic is a measure of the agreement between observers that occurs above that expected by chance. For conditions with a high prevalence, the probability that observers will agree simply by chance is greater than if the condition is less prevalent. Because the prevalence of EIPH, especially low grade, was high in our study, there was a high degree of concordance expected by chance alone and therefore the \( \kappa \) statistics are a conservative estimate of the agreement between observers.

A more useful evaluation of the agreement between observers is provided by the proportion of observations in which 2 or more observers agree and the extent to which they disagree. From a practical viewpoint, differences of 1 EIPH grade are likely of less importance than are differences of 2 or more grades. Observers in our study differed by 1 grade or less in > 99% of examinations, indicating excellent concordance between observers using this grading system.

The grading system was modeled on described systems. The earlier system assigned horses to 1 of 4 or 5 grades including 0, whereas our system assigned horses to 1 of 5 grades. We developed the grading system when it became apparent after examination of several hundred horses that 4 grades of severity of hemorrhage (in addition to lack of hemorrhage) could readily be discerned. Although the performance of the previously described grading systems has not been reported, the high concordance between observers in our study validates the use of our grading system in situations in which there is more than 1 observer.

It is necessary to have a reliable system to evaluate the potential association of EIPH with performance, the role of putative risk factors in the development of EIPH, and the efficacy of therapeutic or prophylactic interventions, such as administration of furosemide or application of nasal dilator strips. Recently, quantification of hemorrhage has been attempted through measurement of RBC count in fluid obtained by use of bronchoalveolar lavage. Although that technique yields numerical results in the form of RBCs per milliliter of lavage fluid and RBC counts vary widely among horses, the relationship between RBC count and severity of hemorrhage has not been determined. There are concerns that because that technique samples only a small portion of 1 lung and hemorrhage may be localized or occur from predominantly 1 lung, the RBC count in lavage fluid may not provide an accurate indication of overall severity of hemorrhage. Conversely, because the technique obtains samples from the distal airways, as opposed to observation of more rostral airways during endoscopic examination, the lavage technique may detect and quantify hemorrhage that is not apparent on endoscopic examination. Finally, collection of bronchoalveolar lavage fluid is invasive, requires sedation of the horse, and often requires administration of local anesthetic solutions into the airways. These requirements render the technique impractical for screening large numbers of horses in a short period, such as may occur during field studies, and for use in horses in competition for which there are restrictions on use of medications. The excellent interobserver reliability of the grading system described in this study provides a tool for grading severity of EIPH in Thoroughbred racehorses. Because the grading system is based on tracheobronchoscopic examination of the airways, which is an examination that can be rapidly and easily performed on unsedated horses after racing, it has the potential to be a valuable tool in the investigation of EIPH.

References