Observer variation in visual assessment of forelimb horseshoe characteristics on Thoroughbred racehorses

Diane K. Gross, DVM, PhD; Susan M. Stover, DVM, PhD; Ashley E. Hill, DVM, MPVM, PhD; Ian A. Gardner, BVSc, MPVM, PhD

Objective—To assess the accuracy and reliability of a visual method of evaluating horseshoe characteristics.

Animals—1,199 Thoroughbred racehorses.

Procedure—Characteristics of 1 forelimb horseshoe were visually assessed on horses immediately prior to racing by 5 field observers at 5 major racetracks in California. Characteristics evaluated included horseshoe type; toe grab height; and the presence of a rim, pad, and heel traction devices. Sensitivity and specificity for observer assessment of horseshoe characteristics were calculated by comparing observer assessments to a postmortem laboratory standard for horses that died within 48 hours of a race. Intraobserver agreement was assessed in a subset of horses by comparing horseshoe observations made before and after the horse's race. Interobserver agreement was evaluated by comparing horseshoe assessment among observers who examined the same subset of horses prior to racing on select days.

Results—The sensitivity and specificity of this visual method of evaluating horseshoe characteristics were good and ranged from 0.75 to 1 and 0.67 to 1, respectively. Agreement beyond chance (weighted kappa values) between observers and the laboratory standard for toe grab height was fair (0.60 to 0.62). Intraobserver and interobserver agreements (kappa values) were high (0.86 to 0.99 and 0.71 to 1, respectively).

Conclusions and Clinical Relevance—Visual observation of horseshoes can be a feasible and reproducible method for assessing horseshoe characteristics prospectively in a large cohort of horses under racing conditions. (*Am J Vet Res* 2004;65:1674–1679)

Musculoskeletal injuries are an important threat to the health and welfare of Thoroughbred racehorses. The occurrence of musculoskeletal injuries has been associated with many factors including demographics, conformation, exercise intensity, and race- and racetrack-related factors.[1-17] Horseshoe characteristics have been implicated as an important contributor to the development of musculoskeletal injuries in Thoroughbred racehorses; however, studies examining toe grabs, rims, and heel traction devices as risk factors for injury have yielded inconsistent results.[1,2,10] Multiple other factors, such as age, race-related factors, racetrack-related factors, and exercise intensity, have also been shown to affect the risk of injury in racehorses. However, their relationship to horseshoe characteristics and therefore their role as potential confounders of the relationship between horseshoe characteristics and musculoskeletal injuries in horses has not been fully evaluated.

To clarify the relationship between horseshoe characteristics and musculoskeletal injuries in Thoroughbred racehorses while controlling for possible covariates, an approach was needed that would allow the comparison of horseshoe characteristics of horses that developed fatal musculoskeletal injuries to those of horseshoes of their surviving competitors. In California, musculoskeletal injuries occur < 2 times out of every 1,000 times a Thoroughbred racehorse starts a race.[8] Therefore, a large cohort of Thoroughbred racehorses would need to be followed prospectively for a long period to obtain the statistical power needed to determine the role of horseshoe characteristics and these potential confounding factors on lameness and musculoskeletal injuries. For this reason, we needed a method of horseshoe assessment that was rapid, noninvasive, economical, feasible to perform at the racetrack under normal operating conditions, and most importantly, not dangerous to the health and welfare of the horses or field personnel. The procedure also had to be acceptable to the racetrack personnel, the racehorse owners and trainers, and the California Horse Racing Board.

Numerous methods for assessing horseshoe characteristics and various tools to accurately measure the height of toe grabs were considered. The use of digital photography and hoof casts to make permanent

Received November 24, 2003.
Accepted April 5, 2004.
From the J. D. Wheat Veterinary Orthopedic Research Laboratory, Department of Anatomy, Physiology, and Cell Biology (Gross, Stover, Hill) and the Department of Medicine and Epidemiology (Gardner), School of Veterinary Medicine, University of California, Davis, CA 95616.
Supported by the National Research Institute (NRI) Competitive Grants Program USDA award No. 2002-02169; the Dolly Green Endowment; the Thoroughbred Owners of California; the California Thoroughbred Trainers Association; the California Horse Racing Board; and the Center for Equine Health, University of California, Davis. Funds provided by the Oak Tree Racing Association, the State of California parimutual fund, and contributions by private donors.
The authors thank A. C. Valenzuela Jr, Francisco Duran, Harry Tosten, Jack Hammonds, Charley Dougherty, and Drs. William Bell, Joan Hurley, and Donald Dooley for their technical assistance.
Address correspondence to Dr. Stover.

records at the racetrack that could be evaluated more fully in a laboratory setting were also considered. However, given the working conditions at the racetrack, these methods were deemed impractical by racetrack employees. For this reason, a method for visually assessing horseshoe characteristics by multiple field technicians was developed that allowed the observation of all horses immediately prior to their race at multiple race meets running simultaneously. Horseshoe inspection was performed without disturbing the normal operations at the racetrack with the cooperation of racetrack personnel and the California Horse Racing Board.

To be effective, a technique was required that was reproducible among and within observers, because horseshoe assessments were to be performed during several years at multiple facilities. The purpose of the study reported here was to evaluate the accuracy and precision of visually evaluating horseshoe characteristics by determining the sensitivity and specificity of this method of assessment, as well as the intraobserver and interobserver agreement.

Materials and Methods

Data collection—Horseshoes of Thoroughbred racehorses were examined immediately before racing at 5 Thoroughbred racetracks in California (northern California: Bay Meadows and Golden Gate Fields; southern California: Del Mar, Hollywood Park, and Santa Anita Park). The characteristics of 1 forelimb horseshoe from each racehorse were recorded from June 17, 2000, to November 3, 2002. However, no information was recorded during racing at fair meetings. During the first year (June 2000 to September 2001), horseshoe information was only recorded for horses racing on dirt surfaces. However, from September 2001 to the end of the study, information was recorded for horses racing on dirt and turf surfaces.

Trained field technicians visually observed 1 forelimb horseshoe of each racehorse immediately before the horse raced. This was done during the farrier inspection required of all horses racing in California. A standardized form[18] was used to record pertinent horseshoe information including the type of horseshoe (eg, none, aluminum racing plate, steel training plate, commercial racing plate,[a] commercial horseshoe,[b] glue-on horseshoes) and whether devices such as a horseshoe pad, rim, heel stickers, jar caulks, heel blocks, or special nails were used.[19] The presence of any toe grab was recorded, and the distance that the toe grab extended from the ground surface of the horseshoe was estimated and used to categorize each toe grab as follows: none = no toe grab; very low = toe grab height of > 0 to ≤ 2 mm; low = toe grab height of > 2 to ≤ 4 mm; regular = toe grab height of > 4 to ≤ 6 mm; and high = toe grab height of > 6 to ≤ 8 mm.

To facilitate future analysis, the categorization of toe grabs in horseshoes with a rim was repeated to reflect the distance measured distally from the ground surface of the rim. Therefore, a horseshoe with a rim that was level around the toe was considered to be in the none toe grab category and a horseshoe with a toe grab that extended 2 mm distally from the ground surface of the rim was considered to be in the very low toe grab category. Also, commercial racing plates[a] with heel traction devices were considered to be in the low toe grab category.

Data collected at the time of the horseshoe examination included the following: racetrack, observer name, date, race number, horse number, horseshoe characteristics for each horse, and name of the horse assigned the number 1 position in each race. The names of the remaining horses were obtained from the official racing program or a commercial data base.[c]

Field observers—Five field observers were used to record horseshoe information. One field observer (observer 1) recorded information for the southern California racetracks for the entire project, and 4 field observers (observers 2, 3, 4, and 5) recorded information for northern California racetracks. One observer (observer 2) recorded information from October 11, 2000, to November 3, 2001; 2 observers (observers 3 and 4) recorded information on separate days between November 8, 2001, and March 31, 2002, and 1 observer (observer 5) recorded information from March 23, 2002, to November 3, 2002. Field observers were either racetrack personnel or had prior equine and veterinary medical experience, but none had previous experience with shoeing racehorses. Observers were trained by 2 of the authors (DKG and SMS) in the identification of horseshoe characteristics via a horseshoe reference board and a handheld standard toe grab height template,[2] which was provided to each observer for the duration of the study. The horseshoe board was used for visual reference and included all of the horseshoes and horseshoe devices of interest to the investigation. An aluminum bar stepwedge of 10 cm in length was used as a handheld standard template. This bar was 8 mm thick at 1 end, and the thickness increased in steps of 2 mm every 2 cm along the length of the bar. The thickness of the bar was made to reflect the thickness of an unworn standard racing plate with no toe grab (8 mm in thickness) and a horseshoe with a very low (10 mm in thickness), low (12 mm in thickness), regular (14 mm in thickness), or high (16 mm in thickness) toe grab. The field observers were instructed on horseshoe characteristics and data collection protocol at the racetracks for 1 to 2 days and were supervised when collecting data during racing. Observers were also encouraged to practice on horses that were not presented for racing to calibrate their visual toe grab height measurements with the standard template. This allowed time for a more detailed examination of the horseshoe than could be obtained during the mandatory prerace examination. Observers recorded information for 2 weeks before their observations were included in project data. To reduce observer variation, the field observers continued to calibrate their visual observations by reference to the horseshoe board and by weekly measurement of the height of at least 1 toe grab with a handheld standard toe grab height template.[2]

Data processing—Data were entered into a computerized spreadsheet[d] and checked for errors by comparison of computer entries with original records and alternative data sources where appropriate.[c,e,f] If a discrepancy was found between the data collected from the race program and that obtained through commercial databases, the commercial data were used.

Assessment of intraobserver and interobserver variation and observer accuracy—Intraobserver variation was assessed beginning in April 2001 by having field observers reexamine the forelimb horseshoes of a subset of horses on the same day as the prerace horseshoe examination. This began once the observers had evaluated horses for at least 10 weeks. Once to twice weekly, the horseshoes of up to 5 horses/d were reexamined after their race. Horseshoe type, toe grab height, pad, rim, jar caulk, heel block, and heel sticker information were recorded. The horses reevaluated were those selected for routine drug testing following their race. These examinations were made without reference to the previous findings of the observer, and the results were recorded separately.

Interobserver variation was assessed by having 2 to 3 field observers simultaneously but independently examine the horses racing on 2 days in northern California (observers 1 and 2 on May 5, 2001; observers 1, 3, and 4 on April 7, 2002) and 3 days in southern California (observers 1 and 2 on May 25, 2001; observers 1 and 3 on April 17, 2002; observers 1 and 5 on April 19, 2002). Because the field observers from northern California were working at different times, comparisons between observers in northern California were not possible.

Observer sensitivity, specificity, and agreement in the assessment of horseshoe characteristics were determined by comparison with findings on forelimb horseshoes from hooves examined in the J. D. Wheat Veterinary Orthopedic Research Laboratory (University of California, Davis, Calif). These hooves were from horses that died or were euthanatized within 48 hours of a field observation and necropsied through the California Horse Racing Board Postmortem Program.[11] The characteristics of the horseshoes were recorded with the same criteria as those used by the field observers. The height of any toe grab was evaluated by visually comparing the horseshoe on the cadaveric foot to a standard template.[10,18] If the toe grab was unevenly worn, the maximal height was used. Although the standard template used in the laboratory was designed to reflect the thickness of the standard racing horseshoe (ie, racing plate with no toe grab [8 mm in thickness] and racing plates with very low [10 mm in thickness], low [12 mm in thickness], regular [14 mm in thickness], and high [16 mm in thickness] toe grabs), actual thickness of the templates[g] used for laboratory measurements varied slightly from this design (ie, templates were categorized as none [ie, equivalent to no toe grab, 8 mm in thickness], very low [10 mm in thickness], low [11.5 to 11.8 mm in thickness], regular [14.4 mm in thickness], and high [15.8 to 16.1 mm in thickness]).

Data analysis—Sensitivity, specificity, and kappa values (95% **confidence intervals [CIs]**)[h] for observer assessment of horseshoe characteristics were calculated to evaluate accuracy and agreement of observers. When the kappa values were 1, it was not possible to calculate 95% CI. Kappa statistics were weighted with Cicchetti-Allison weights, which assign greater weights for those disagreements that were further from the diagonal (agreement) on an ordinal scale table and lesser weights to disagreements close to the diagonal. Overall and simple kappa values were used to evaluate differences in agreement among observers. Differences between paired toe grab heights were evaluated by means of a signed rank test.[i] Wherever necessary, categoric data were collapsed to ensure adequate numbers of horses per category and further evaluate accuracy and agreement between different levels of categoric data. Information for an individual horse may have been included from multiple days if the horse was reevaluated > 1 time. Values are not reported for characteristics when data were sparse. Values of $P < 0.05$ were considered significant.

Results

Intraobserver agreement—During the study, 64,528 horseshoes were evaluated on 11,336 horses. Horseshoes from 811 horses (499 from southern California and 312 from northern California) were evaluated by observers (observers 1, 2, 3, 5) before the race and then again after the race to evaluate intraobserver agreement **(Table 1)**. However, no horses were reevaluated by observer 4. On average, no difference was found in the height of toe grabs recorded before and after the race. The mean (± SEM) difference in toe grab height measurements was -0.001 (± 0.006) categoric levels ($P = 1.0$). Kappa (95% CI) value, which measured intraobserver agreement beyond chance, revealed that intraobserver agreements were good to excellent and ranged from 0.74 to 1 for different horseshoe characteristics for each observer. For example, when comparing the horseshoe type recorded during the prerace inspection with the type recorded during the follow-up examination, observer 2 had a kappa value of 0.74, indicative of at least moderate agreement between the 2 observation times because a kappa value of 0 indicates no agreement beyond chance, whereas 1 indicates complete agreement.[20] For all observers, kappa values ranged from 0.86 to 0.99 and no significant differences were found between observer values.

Interobserver agreement—For 326 horses evaluated at the same time by the observer from southern California (observer 1) and the observers from northern California (observers 2, 3, 4, 5), interob-

Table 1—Intraobserver* agreement beyond chance for evaluation of horseshoe characteristics for Thoroughbred racehorses in California.

Horseshoe characteristics	Combined (n = 811)		Observer 1 (499)		Observer 2 (159)		Observer 5 (128)		Observer 3 (25)	
	kappa	95% CI	kappa	95% CI	kappa	95% CI	kappa	95% CI	kappa	95% CI
Shoe type	0.93	0.85–1	1	NA	0.74	0.40–1	0.96	0.87–1	1	NA
Toe grab (by 4 categories)†	0.99	0.98–1	0.99	0.98–1	0.98	0.96–1	1	NA	0.92	0.76–1
Rims	0.98	0.96–1	0.99	0.98–1	0.95	0.89–1	1	NA	0.88	0.66–1
Pad	0.98	0.96–0.99	0.99	0.97–1	0.91	0.83–1	1	NA	0.86	0.61–1
Caulk/Block/Sticker	0.86	0.58–1	ND	ND	0.85	0.57–1	ND	ND	ND	ND

*Data from observer 4 not evaluated. †Weighted kappa used.
CI = Confidence interval. NA = Not applicable. ND = Not determined.

Table 2—Interobserver agreement beyond chance between observers from northern and southern California comparing toe grab height for Thoroughbred racehorses.

Observers	Toe grab height (categories)	Southern California (No. of horses)* None	Very low	Low	Regular or high	Totals
Northern	None	33	39	3	3	78
California	Very low	20	49	3	0	72
(No. of horses)†	Low	0	5	62	5	72
	Regular or high	1	3	21	79	104
	Totals	**54**	**96**	**89**	**87**	**326**

*Southern California includes data obtained by field observer 1. †Northern California includes combined data obtained by field observers 2, 3, 4, and 5.

server agreement beyond chance for toe grab height measurements was good with a weighted kappa value of 0.71 (95% CI, 0.66 to 0.76) and a difference of only 1 toe grab height categoric level was recorded in 90% of all disagreements (**Table 2**). Most disagreements (82%) occurred between the none to very low toe grab categories and between the low to regular or high toe grab categories. However, on average, the interobserver toe grab height was not different between the observer from southern California and the observers from northern California. The mean (± SEM) difference in measurements of toe grab height was –0.03 (± 0.04) categoric levels ($P = 0.48$). The kappa value for rims was 0.80 (95% CI, 0.70 to 0.89), 0.90 (95% CI, 0.84 to 0.95) for pads, and 1 for horseshoe type. Interobserver agreement was not evaluated for heel traction devices as a result of small sample size. The observer from southern California recorded heel traction devices on 3 horses, but observers from northern California recorded none.

The interobserver agreement beyond chance for toe grab height measurements between observer 1 and each northern California observer (observers 2 [n = 98], 3 [104], 4 [57], and 5 [69]) was fair to good with a range of weighted kappa from 0.64 (95% CI, 0.52 to 0.76) for observer 5 to 0.71 (95% CI, 0.69 to 0.82) for observer 4. The kappa value for rims ranged from 0.56 (95% CI, 0.18 to 0.95) for observer 4 to 0.90 (95% CI, 0.80 to 1) for observer 2. For pads, interobserver agreement beyond chance ranged from a kappa value of 0.84 (95% CI, 0.62 to 1) for observer 4 to 0.92 (95% CI, 0.84 to 1) for observer 3. All 4 observers were in complete agreement with observer 1 on horseshoe type, with a kappa value of 1.

Comparison to laboratory standard—Horseshoes from 31 horses from southern California (observer 1) and 31 from northern California (observers 2 [n = 15], 3 [6], 4 [4], and 5 [6]) were available for postmortem examination. With the exception of horseshoe type and the presence of a toe grab, the overall sensitivity and specificity for observer assessment of all horseshoe characteristics in our study was 0.83 (**Table 3**). However, when toe grab was dichotomized (ie, evaluated as a yes or no category), the overall sensitivity and specificity for observer assessment were only 0.75 and 0.67, respectively. As a result of sparse data, sensitivity, specificity, and kappa values for assessment of horseshoe characteristics for individuals were not reported separately.

Table 3—Sensitivity and specificity of assessment of 5 observers on the horseshoe characteristics for 62 Thoroughbred racehorses in California.

Horseshoe characteristics	Observers All	Southern California†	Northern California*
		Sensitivity	
Shoe type	1	1	1
Toe grab‡	0.75	0.77	0.72
Toe grab§	0.93	0.94	0.91
Rim	0.91	1	0.83
Pad	0.86	0.88	0.75
Heel traction	0.83	NA	0.75
		Specificity	
Shoe type	NA	NA	NA
Toe grab‡	0.67	NA	NA
Toe grab§	0.89	0.87	0.90
Pad	0.95	1	0.93
Heel traction	1	1	1

*Southern California includes data obtained by field observer 1 (n = 31). †Northern California includes combined data obtained by field observers 2, 3, 4, and 5 (n = 31). ‡Yes versus no. §None to very low toe grab categories versus low to regular or high toe grab categories.

NA = Not assessed because of sparse data.

When toe grab height was measured as a 4-category variable during prerace inspection and was compared with horseshoe characteristics observed during postmortem examination (at the veterinary orthopedic research laboratory), the weighted kappa value was 0.6 for northern and southern California (**Table 4**), and no significant differences were found between regions. Individual kappa values for toe grab height for each observer were not calculated because of sparse data. On average, the toe grab height recorded was greater during the postmortem examination than during prerace inspection. Mean (± SEM) difference in measurements of toe grab height between the postmortem examination and prerace inspection was 0.3 (± 0.8) categoric levels ($P = 0.005$).

Table 4—Evaluation of agreement beyond chance between observers and the laboratory standard for toe grab height of 62 Thoroughbred racehorses in California.

		Laboratory standard (No. of horses)				Kappa	
Observers*	Toe grab height (categories)	None	Very low	Low	Regular or High	Weighted	95% CI
Northern California (No. of horses)	None	2	6	1	1	0.61	0.42–0.79
	Very low	0	2	0	0		
	Low	0	0	3	1		
	Regular or high	0	1	2	12		
Southern California (No. of horses)	None	0	7	0	0	0.60	0.42–0.78
	Very low	1	7	2	0		
	Low	0	0	7	1		
	Regular or high	0	1	0	5		
All Observers (No. of horses)	None	2	13	1	1	0.62	0.50–0.75
	Very high	1	9	2	0		
	Low	0	0	10	2		
	Regular or high	0	2	2	17		

*Southern California includes data obtained by field observer 1 (n = 31). Northern California includes combined data obtained by field observers 2, 3, 4, and 5 (n = 31). All observers includes data from all observers in northern and southern California (n = 62).

Discussion

During our study, 5 different observers examined approximately 64,500 horseshoes during the prerace farrier inspection at 5 major racetracks in California. This method of visual observation allowed the rapid assessment of horseshoes worn by all Thoroughbred racehorses without impeding the normal activity at the racetracks. Overall, sensitivity and specificity for observer assessment of horseshoe characteristics (ie, toe grab height, horseshoe type, rim, pad, and heel traction devices) on the basis of 62 observations were good and ranged from 0.75 to 1 and 0.67 to 1, respectively. Intraobserver and interobserver agreement beyond chance were high for all observers. Intraobserver agreement assessed on 811 observations ranged from a kappa value of 0.86 to 0.99, whereas interobserver agreement assessed on 326 observations ranged from a kappa value of 0.71 to 1. Therefore, this method of visual observation of horseshoes was a feasible and consistent method for assessing horseshoe characteristics prospectively in a large cohort of horses.

Careful validation of a subjective measurement device is an important part of clinical research to facilitate accurate interpretation of the results and evaluate the usefulness of a measurement device in future studies. For example, the usefulness of this horseshoe assessment technique can be predicted for future studies that evaluate the effect of toe grab height on the risk of fatal musculoskeletal injury. Results of a previous study by Kane et al[10] in which postmortem findings in Thoroughbred racehorses in California were used, indicated that horses that died of musculoskeletal injuries were 2 times as likely (**odds ratio [OR]**, 2.1; 95% CI, 1.5 to 2.9) to have horseshoes with toe grabs with low, regular, or high height measurements versus horseshoes without toe grabs or toe grabs with very low height measurements, compared with horses that died from other causes.

If horseshoes from horses in active race training that sustained fatal musculoskeletal injuries were compared with horseshoes from racing competitors that did not have a fatal injury, our visual assessment method should be adequate to detect a difference in musculoskeletal injuries when examining toe grab height as a categoric variable. In our study, overall sensitivity and specificity were 0.93 and 0.89, respectively, for observer assessment of horseshoes without toe grabs and horseshoes with toe grabs with very low height measurements, compared with horseshoes with toe grabs with low, regular, or high height measurements. By use of this visual method of toe grab height assessment, a case-control study of 250 horses with musculoskeletal injuries and 750 racehorses that survived their race (racehorses without musculoskeletal injuries) would result in an estimated OR similar to the OR that would have been obtained using the standard method, assuming that incorrect estimation of toe grab height was random (nondifferential) within categories of the risk factor. If the estimated OR of horseshoes with low, regular, or high toe grab height measurements versus horseshoes without toe grabs or horseshoes with very low toe grab height measurements is 2.1 (95% CI, 1.5 to 2.9) times as likely in racehorses with musculoskeletal injuries, compared with racehorses without musculoskeletal injuries, the true OR would only be 2.2 (95% CI, 1.6 to 3.1) given a sensitivity and specificity for observer assessment of horseshoe characteristics of 0.90. Practically, this difference is negligible.

Agreement, sensitivity, and specificity for observer assessment of horseshoe characteristics were often determined with few horses for each observer in our study. Therefore, we were unable to compare some observers for certain uncommon horseshoe characteristics. This was particularly true for estimation of individual observer sensitivities and specificities, where only 4 to 31 horseshoes were available for postmortem examination to evaluate the accuracy of each observer. For sensitivity and specificity determination, if a minimum of 3 horseshoes was not available within each horseshoe category recorded by either the observer or at the veterinary orthopedic research laboratory, the test was not evaluated. This occurred for 44% (22/50) of individual sensitivity and specificity measurements.

Differences between measurements on postmortem examination and prerace inspection were detected, and field observers most commonly underestimated the toe grab height, compared with the laboratory standard. Evaluating toe grab height on an ordinal scale allowed for discrimination of small versus large measurement differences. Although the sensitivity and specificity for all observers for any toe grab were 0.75 and 0.67, respectively, the weighted kappa was fair to good (0.62; 95% CI, 0.50 to 0.75). Although 39% (24/62) of the toe grab height field observations disagreed with the laboratory standard, only 17% (4/24) of these disagreements differed by 1 categoric level, and the field observers underestimated the height of toe grabs, compared with the laboratory standard, in 79% (19/24) of the disagreements. Overall, sensitivity and specificity of observers for visual observation of horseshoe characteristics were good. However, individual observer sensitivity, specificity, and kappa values had more variation. The weighted kappa for toe grab height was 0.6 when comparing northern and southern California observers were compared with the laboratory postmortem standard, whereas sensitivities and specificities ranged from 0.67 to 1. Because of the small sample sizes for individual observers, even infrequent differences between postmortem examination and field examination data could result in large changes in sensitivity and specificity values. Also, differences in toe grab height measurement may have been attributed to a difference in the method used to evaluate horseshoes with uneven wear. In the postmortem evaluation, the maximal height measured anywhere on the toe grab was recorded. Visual assessment of toe grab height in the field may not have been able to detect small differences in toe grab height in unevenly worn horseshoes, where the mean toe grab height was estimated and recorded. Unfortunately, the amount of wear on the horseshoe was not recorded to account for this difference.

Visual evaluation of horseshoes of racehorses prior to racing allowed observers to examine horses in a natural setting without interfering with racing. This made it possible to follow a large cohort of horses in multiple locations, which is necessary to obtain a sample size adequate to study risk factors and potential confounders for a rare event such as catastrophic musculoskeletal injury. Visual observation of horseshoes was a practical, rapid, and low cost method of assessing horseshoe characteristics of Thoroughbred racehorses at multiple racetracks that produced reliable results. However, in a smaller scale study where conditions are more conducive to the manipulation of hooves, measurement of toe grab height likely will be more accurate and reproducible than visual assessment.

[a]World Racing Plate, Thoro'Bred Inc, Anaheim, Calif.
[b]Natural Balance Shoe, Thoro'Bred Inc, Anaheim, Calif.
[c]Jockey Club Information Systems, Lexington, Ky.
[d]Microsoft EXCEL 2000, Microsoft Corp, Redmond, Wash.
[e]Bloodstock Information Research Services, Lexington, Ky.
[f]California Horse Racing Board, Sacramento, Calif.
[g]Mitutoyo Corporation, Kawasaki, Japan.
[h]PROC FREQ, SAS Institute Inc, Cary, NC.
[i]PROC UNIVARIATE, SAS Institute Inc, Cary, NC.

References

1. Balch OK, Helman RG, Collier MA. Underrun heels and toe-grab length as possible risk factors for catastrophic musculoskeletal injuries in Oklahoma racehorses, in *Proceedings*. 29th Annu Meet Am Assoc Equine Pract 2001;334–338.
2. Hill AE, Stover SM, Gardner IA, et al. Risk factors for and outcomes of noncatastrophic suspensory apparatus injury in Thoroughbred racehorses. *J Am Vet Med Assoc* 2001;218:1136–1144.
3. Hill T, Carmichael D, Maylin G, et al. Track condition and racing injuries in Thoroughbred horses. *Cornell Vet* 1986;76:361–379.
4. Hill AE, Carpenter TE, Gardner IA, et al. Evaluation of a stochastic Markov-chain model for the development of forelimb injuries in Thoroughbred racehorses. *Am J Vet Res* 2003;64:328–337.
5. Mohammed HO, Hill T, Lowe J. Risk factors associated with injuries in Thoroughbred horses. *Equine Vet J* 1991;23:445–448.
6. Estberg L, Stover SM, Gardner IA, et al. Fatal musculoskeletal injuries incurred during racing and training in Thoroughbreds. *J Am Vet Med Assoc* 1996;208:92–96.
7. Estberg L, Stover SM, Gardner IA, et al. High-speed exercise history and catastrophic racing fracture in Thoroughbreds. *Am J Vet Res* 1996;57:1549–1555.
8. Estberg L, Stover SM, Gardner IA, et al. Relationship between race start characteristics and risk of catastrophic injury in Thoroughbreds: 78 cases (1992). *J Am Vet Med Assoc* 1998;212:544–549.
9. Hernandez J, Hawkins DL, Scollay MC. Race-start characteristics and risk of catastrophic musculoskeletal injury in Thoroughbred racehorses. *J Am Vet Med Assoc* 2001;218:83–86.
10. Kane AJ, Stover SM, Gardner IA, et al. Horseshoe characteristics as possible risk factors for fatal musculoskeletal injury of Thoroughbred racehorses. *Am J Vet Res* 1996;57:1147–1152.
11. Kane AJ, Stover SM, Gardner IA, et al. Hoof size, shape, and balance as possible risk factors for catastrophic musculoskeletal injury of Thoroughbred racehorses. *Am J Vet Res* 1998; 59:1545–1552.
12. Kobluk CN, Robinson RA, Gordon BJ, et al. The effect of conformation and shoeing: a cohort study of 95 Thoroughbred racehorses, in *Proceedings*. 35th Annu Meet Am Assoc Equine Pract 1990; 259–274.
13. Rooney JR. The relationship of length of race to fatigue and lameness in Thoroughbred racehorses. *J Equine Vet Sci* 1982;2:98–101.
14. Peloso JG, Mundy GD, Cohen ND. Prevalence of, and factors associated with, musculoskeletal racing injuries of Thoroughbreds. *J Am Vet Med Assoc* 1994;204:620–626.
15. Cohen ND, Berry SM, Peloso JG, et al. Association of high-speed exercise with racing injury in Thoroughbreds. *J Am Vet Med Assoc* 2000;216:1273–1278.
16. Carrier TK, Estberg L, Stover SM, et al. Association between long periods without high-speed workouts and risk of complete humeral or pelvic fracture in Thoroughbred racehorses: 54 cases (1991–1994). *J Am Vet Med Assoc* 1998;212:1582–1587.
17. Lindner A, Dingerkus A. Incidence of training failure among Thoroughbred horses at Cologne, Germany. *Prev Vet Med* 1993;16:85–94.
18. Gross DK, Stover SM, Hill AE, et al. Forelimb horseshoe characteristics of Thoroughbreds racing on dirt surfaces. *Am J Vet Res* 2004;65:1021–1030.
19. Stashak TS. Classification of horseshoes and horseshoe nails. In: Stashak TS, ed. *Adams' lameness in horses*. 4th ed. Philadelphia: Lea & Febiger, 1987;786–806.
20. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–174.