A COMPARATIVE STUDY OF FULL HINDLIMB FLEXION IN HORSES: 5 VERSUS 60 SECONDS

by

AMANDA RAE ARMENTROUT

B.S., University of Nevada, 2002

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Department of Clinical Sciences
College of Veterinary Medicine

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2011

Approved by:

Major Professor
Warren L. Beard, DVM, MS, DACVS

Abstract

The flexion test is routinely used in lameness and pre-purchase examinations. There is no accepted standard for duration of flexion or evidence that interpretation of results would differ with different durations of flexion. We hypothesized there would be no difference in interpretation of full hindlimb flexion for 5 or 60 seconds. Video recordings of lameness examinations of 34 client owned horses were performed that included: baseline lameness, upper hindlimb flexion for 60 seconds, and flexion of the same leg for 5 seconds. Videos were edited to blind reviewers to the hypothesis being tested. The baseline lameness video from each horse was paired with each flexion to make 2 pairs of videos for each case. Twenty video pairs were repeated to assess intra-observer repeatability. Fifteen experienced clinicians reviewed the videos and graded the response to flexion as either positive or negative. Potential associations between the duration of flexion and the likelihood of a positive flexion test were evaluated using generalized linear mixed models. A kappa value was calculated to assess the degree of intra-observer agreement on the repeated videos. Full hindlimb flexion of 60 seconds was more likely to be called positive than flexion of 5 seconds ($p<0.0001$), with the likelihood of the same interpretation 74% of the time. The first flexion performed was more likely to be called positive than subsequent flexions ($p=0.029$). Intra-assessor agreement averaged 75% with $\kappa=0.49$. Full hindlimb flexion of a horse for 5 seconds did not yield the same result as 60 seconds.

Table of Contents

Acknowledgements

I would like to extend special thanks to our esteemed colleagues who agreed to serve as our reviewers: Jerry Black, DVM, Tom Bohanon, DVM, MS, DACVS, Kelly Farnsworth, DVM, MS, DACVS, Steve Hance, DVM, DACVS, Brad Jackman, DVM, MS, DACVS, Andis Kaneps, DVM, PhD, DACVS, David G. MacDonald, DVM, MVSc, DACVS, Omar Maher, DVM, DACVS, Stephanie Mathis, DVM, DACVS, Brook Mayberry, DVM, PhD, Shane M. Miller, DVM, DACVS, William Neal, Jr., DVM, Al Ruggles, DVM, DACVS, Wes Sutter, DVM, MS, DACVS, and Wyatt Winchell, DVM, MS, DACVS. I would also like to thank Dr. Marc Holmes and Dr. Hugh Townsend for their help with the design of the project.

A very special thank you belongs to Dr. Warren L. Beard for his mentorship and support during this project as well as a thank you to Dr. Brad White for his help with the data analysis.

Chapter 1 - Introduction and Materials and Methods

Introduction

Lameness has the highest annual incidence of all equine medical problems (Kane and National Animal Health Monitoring System (U.S.) 2000) and is the most important health issue to horse owners and trainers (Kaneene *et al.* 1997). Lameness evaluations are a part of everyday life for the equine practitioner. Flexion tests are an integral part of the lameness examination and are used to exacerbate the baseline lameness as well as reveal a hidden lameness (Dyson 1998; Ross 2003).

A lameness exam is unequivocally subjective and numerous studies have illuminated the lack of agreement among observers (Arkell *et al.* 2006; Fuller *et al.* 2006; Hewetson *et al.* 2006; Keegan *et al.* 2010; Keegan *et al.* 1998; Ramey 1997). Interpretation of a flexion test is also subjective and each practitioner interprets their result based on the signalment and use of the horse, presence or absence of other corroborating physical exam findings, degree of the lameness accentuation, and knowledge of the amount of force applied during the flexion. Although one examiner's results are likely not repeatable by others (Keg *et al.* 1997a; Keg *et al.* 1997b; Verschooten and Verbeeck 1997), one study has demonstrated that intra-observer repeatability of the exam by a single practitioner is good with a reported mean coefficient of variation being only 12% (Keg *et al.* 1997b). There is little agreement between clinicians about the length of time a joint should be flexed and this has prompted investigation into fetlock flexion duration with the goal of standardization. Fetlock flexion of 30 seconds does not always reveal a positive result; whereas flexion of two minutes yields a false positive result (Keg *et al.* 1997a), with the optimal duration of flexion determined to be 60-90 seconds. Similar scrutiny has not been applied to full limb flexion of the hindlimbs. The recommended duration for full hindlimb flexion is purely empirical and ranges from one to two minutes (Dyson 1997; Lacroix 1916; Ross 2003; Stashak 2002).

During a soundness examination it is likely that flexion tests will be performed on multiple joints of all limbs. Flexing multiple joints for 60-90 seconds results in considerable time investment that can be tiring to the practitioner, and occasional problems with patient compliance which can endanger the personal safety of the practitioner. The authors are aware that some practitioners perform flexion tests only for the length of time necessary to obtain maximal flexion. The authors are not aware of any evidence to suggest that short periods of flexion would not result in the same interpretation. If a shorter duration of flexion would result in the same interpretation as a longer duration then the lameness evaluation process could be significantly shortened without reducing the value of the test. The purpose of this study was to determine if practitioners will reach the same interpretation of full limb flexion of the hindlimbs when a flexion test is performed for five or sixty seconds. We hypothesized that there would be no difference in interpretation of full hindlimb flexion for five or sixty seconds.

Materials and Methods:

Horses

Thirty four client-owned horses were presented to Kansas State University and evaluated for lameness. Horses ranged in age from 4 to 25 years old with a mean age of 9.4 years (median age of 9). There were 14 mares and 20 geldings in the study population, and breeds included were 25 Quarter Horses, 4 Thoroughbreds, 3 Warmbloods, 1 pony, and 1 Arabian. Horses that were selected for the study were solid in color and bore no obvious markings or brands. Horses were excluded if they were fractious or would not trot in a straight line for video recording. Horses were selected such that approximately one third of the horses had no obvious response to full hindlimb flexion, one third of the horses had a mild to moderate response to flexion, and one third of the horses had a marked response to flexion. This response to flexion was the clinical interpretation made by the authors at the time of the lameness examination. These horses ranged from clinically normal (no lameness evident on baseline examination) to markedly lame. The described procedures were approved by the Kansas State University Institutional Animal Care and Use Committee.

Pilot Study

An unblinded pilot study was conducted by the authors for six months prior to the start of this study. During routine lameness examinations, the hindlimbs of horses were flexed by the examiners for sixty seconds as well for five seconds. The duration of the first flexion was arbitrarily chosen and the duration between repeated flexion tests was 1-3 minutes. The authors felt that they interpreted the response to flexion the same for both time periods, and a cumulative effect of flexing a limb was not observed. Additionally, for 3 months prior to the start of this study, routine clinical lameness examinations with flexion tests were videotaped and those tapes were edited as described below. This set of blinded videos was shown to a single author for interpretation of response to flexion, and he consistently interpreted the results of the two durations of flexion the same.

Study Design

Each horse had a lameness evaluation performed. The baseline gait was recorded by digital video with the horse trotting away in a straight line for 35 meters. Video recordings were then made of the flexion tests performed on a single hindlimb. Full limb flexion of one limb was performed and each limb underwent two flexion tests; with durations of five seconds and sixty seconds respectively. The order of flexion tests was determined by random assignment. Time between flexion tests ranged from one to three minutes. Full hindlimb flexion was performed with the clinician standing to the side of the horse and lifting the hindlimb until the metatarsus was parallel to the ground and the tarsal region was in maximal flexion (Figure 1). The limb was held by the metatarsus for the pre-determined time.

Videotaping

Video recordings were made of each horse. The horses were trotted away from the camera in a straight line on a hard surface. In an attempt to maintain consistency and to blind reviewers, the recordings were made at the same site with identical starting and ending points for each horse. The camera was placed in the same location for each exam: three meters directly behind the horse (Figure 2). There were two technicians that were designated to trot the horses who were similar in size and hair color. They wore standardized clothing, and the same technician trotted each individual horse for the entirety of the examination. The same individual performed all the flexion examinations, and was dressed in standardized clothing.

Editing

Each horse had 3 videos associated with the examination: a baseline lameness video, a five second flexion video, and a sixty second flexion video. The flexion videos were edited to show only the last two seconds of the flexion and subsequent trotting. The duration of each video was 10-14 seconds. The videos were paired so that every horse had a baseline lameness video paired with each of the flexion videos of the single limb, resulting in two video sets per horse (baseline with a five second flexion and baseline with a sixty second flexion). A computer spreadsheet program (Microsoft Excel 2007, Microsoft Corporation Redmond, WA) was used to randomly select 10 of the 34 cases to repeat. The 10 cases had both sets of videos repeated. All videos including the repeated cases were randomly sorted. This generated a list of 88 video sets in random order which were then named by that order (Case 1-88). These 88 cases were made up of 34 horses with two video sets totaling 68 videos and the 10 repeated horses with two video sets totaling an additional 20 videos. The videos were then made into a DVD, with a unique case number associated with each pair of videos.

Reviewers

Twenty one reviewers agreed to participate in the study; however only 15 reviewers ultimately completed the study. Reviewers were known to the authors and considered to be experienced equine practitioners with a range of eight to thiry nine years in practice and a mean of 20 years. They were employed either in private practice or University Teaching Hospitals. Twelve of the reviewers were Diplomates of the American College of Veterinary Surgeons and two reviewers had a PhD. The reviewers were blinded to the hypothesis and purpose of the study. They were instructed to watch the baseline lameness video first and then watch the video of the flexion exam. They were allowed to watch the baseline and flexion videos as many times as they deemed necessary to make a judgment. They were asked to interpret the result of the flexion exam as either positive or negative based on their own criteria. No instructions or guidelines were given as to what the authors interpreted as a positive or negative response to flexion.

Statistical Analysis

Data from reviewers' score cards were entered into a computer spreadsheet program (Microsoft Excel 2007, Microsoft Corporation, Redmond, WA) and imported into a commercial software (SAS 9.2, SAS Institute, Cary NC) for statistical analyses. A data set was created that did not include the data from repeated videos to evaluate inter-observer variability and the potential effect of flexion duration. Potential associations between the duration of flexion (five vs. sixty seconds) and the individual reviewer, with the probability of scoring a flexion test positive, were evaluated using a generalized linear mixed model (proc GLIMMIX, SAS Institute). The individual horse was included as a random effect in the model to account for the lack of independence between the five and sixty second flexion tests performed on the same horse. The test performed first (five or sixty seconds) was also included in the model to account for potential outcome variation related to test order. A P value of <0.05 was chosen for statistical significance.

Intra-observer agreement was calculated from the reviewers' responses from the videos of the 10 horses that were repeated (10 horses, two flexion videos repeated twice). A data set was created pairing the repeated responses and a kappa statistic was calculated (SAS 9.2, SAS Institute, Cary NC) to compare the agreement beyond chance. Additionally, simple descriptive statistics were used to determine how often the reviewers agreed with themselves on identical videos.

Figures

Figure 1.1 Flexion of the hindlimb

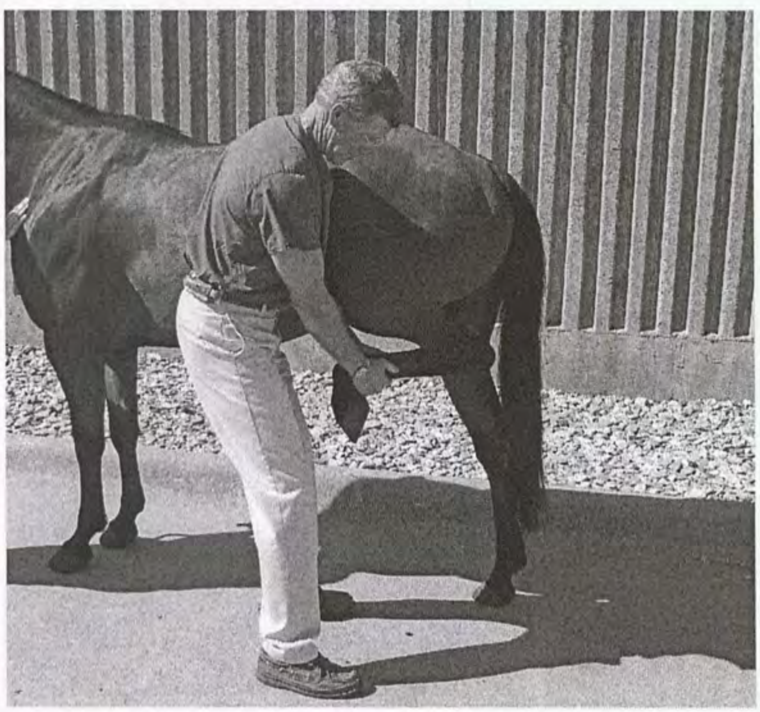

Figure 1.2 Camera directly behind horse

Chapter 2 - Results and Discussion

Results

The probability the flexion test was positive was significantly ($P<0.05$) associated with reviewer, order of the flexion test (five vs. sixty seconds), and flexion length. When evaluating the probability of a positive flexion test, there was no significant ($P = 0.17$) interaction between flexion length and reviewer; therefore, only main effect results are reported. Individual reviewers interpreted response to flexion differently, with a range of 21%-81% of all flexion tests scored positive. Horses with full hindlimb flexion of sixty seconds had a higher probability of being called positive (54% +/- 10) than flexion of five seconds (36% +/- 9, $p<0.0001$). Individual reviewers interpreted the results of five seconds of flexion the same as sixty seconds on average 74% of the time (range 56-90%). The first flexion performed was more likely to be called positive (49% +/-10) than the subsequent flexion (40% +/- 9, $p=0.029$).

Intra-observer agreement as concluded by a kappa statistic was acceptable ($\kappa = 0.49$, 95% CI 0.39-0.59) when examining the results of reviewers interpretation of flexions for the repeated 10 horses. Reviewers agreed with themselves (interpreting a single flexion test as positive or negative) on the repeated 10 horses an average of 75% of the time with a range of 55-90% for each of the individual reviewers.

Discussion

There are many aspects concerning the diagnosis or detection of lameness and in particular concerning the interpretation of full hindlimb flexion testing that were not addressed by this study design. These factors include but are not limited to the source of pain, the influence of force of flexion, and the value of flexion tests in the detection of lameness. It is both a strength, and possibly a weakness, that we focused on a single input variable (duration of flexion) and tried to eliminate any factor that would lead to systematic bias. For example, in a more representative clinical situation, the clinician would be able to vary their position to look at the horse from different angles. They would repeat the test if there was a marginal result, and they would not consider the flexion test result in isolation of other clinical data such as history and the results of other tests. However by concentrating on this one aspect of a single test we can produce the best evidence on the effect of the time the joints are held in flexion, since the objective of this study was simply to determine if there was a difference between the amount of lameness detected when a hindlimb was held in flexion for five seconds as opposed to sixty seconds.

There was great variability between reviewers in interpretation of flexion tests. Individual reviewers graded the response to flexion as positive from as few as 21% to as high as 81% of all flexion tests evaluated. This is not unexpected based on prior literature addressing inter-observer agreement in a clinical examination (Arkell *et al.* 2006; Fuller *et al.* 2006; Hewetson *et al.* 2006; Keegan 2007; Keegan *et al.* 1998). It may also represent the varied background of the reviewers and the diverse breeds of horses they work on. For example, one practitioner works almost solely with western performance horses while another works almost solely with sport horses. It is likely that some reviewers were overly critical and called any minor change in gait a positive response while others may have disregarded minor changes in gait(such a positive response for only a few steps) as a negative. The lack of uniformity in how practitioners judge the response to flexion in no way diminishes the findings of this investigation because inter-observer agreement was not at issue. Intra-observer agreement was the only consideration. Specifically, does an individual practitioner reach the same conclusion when a limb is flexed for either five seconds or sixty seconds.

Overall, reviewers were more likely to judge flexion of sixty seconds positive than flexion of five seconds (54% vs. 36%). This would appear to provide the simple answer to the question that the study attempted to address. Interestingly, individual reviewers interpreted the results of flexion for five seconds the same as flexion for sixty seconds on average 74% of the time, with some reviewers judging response to flexion following five seconds and sixty seconds the same up to 90% of the time. Flexing the limb for a shorter duration may be useful for practitioners who have good agreement on interpretation of flexions for five and sixty seconds. Each veterinarian should perform their own trial to determine if shorter duration of flexion is useful in their circumstances. What is not clear is whether five seconds underestimates the true incidence of positive result or whether flexion of sixty seconds overestimates the incidence. It is likely that both scenarios may occur depending on the individual horse and observer. This question is not answerable in the absence of an accepted standard for what constitutes a positive result. The results of our study highlight the lack of standard interpretation of a positive flexion test, because the range of positive responses ranged between 21% and 81% when viewing 68 identical videos.

The first flexion performed on a patient, regardless of duration, was more likely to be called positive than the second flexion (49% versus 40%, respectively). This is in contrast to that which Busschers and van Weeren observed on flexion of the distal aspect of the forelimbs (Busschers and van Weeren 2001). While they determined that flexion of the distal aspect of the forelimbs could be repeated once without affecting the reliability of the flexion exam, they also demonstrated a cumulative effect of flexion tests on scoring of lameness, with repeated flexions interpreted as significantly more positive. The authors cannot explain why the response to repeated flexion of the fetlock is different from full limb flexion of the hindlimbs. The most likely explanation is that the horses are warmed up by the first flexion; however, given the small difference, the possibility of a type 1 statistical error cannot be excluded.

The video recordings of 10 horses were repeated to allow us to assess intra-observer repeatability. Overall, practitioners reached the same interpretation on these repeated videos 75% of the time. The kappa statistic indicates acceptable agreement ($\kappa = 0.49$) on the scale used by Martin and Bonnett (Martin and Bonnett 1987). This scale indicates κ values between 0.3 and 0.5 are acceptable, κ values between 0.5 and 0.7 are good, and $\kappa > 0.7$ is excellent agreement. The intra-observer repeatability of this study is similar to other studies performed on a variety of lameness conditions (Fuller *et al.* 2006; Keegan *et al.* 1998). The intra-observer repeatability in these studies ranged from $\kappa = 0.34$ to 0.79.

It was impossible to completely blind reviewers to the repetition of horses and videos because weather conditions and size and color of the horse cannot be edited out of videos, so some practitioners may have detected they were seeing the same horses as well as the same videos. The attempt at total blinding of our reviewers to hypothesis and treatment group was imperative to prevent the bias that another study observed (Arkell *et al.* 2006). In that study, when reviewers knew that a treatment had been applied (a nerve block), the interpretation of lameness changed from their own blinded interpretation of the same examination. This 'hindsight bias' has been observed in human medicine, in which physicians' knowledge of a likely diagnosis impacted their certainty of a diagnosis (Dawson *et al.* 1988).

We included a larger number of horses (n=34) and reviewers (n=15) than most previous studies (Fuller *et al.* 2006; Keegan *et al.* 2010; Keegan *et al.* 1998). The reviewers were from different backgrounds, in different types of practice, and scattered over a broad geographic area. Due to the large number of reviewers and horses in this study, inferences drawn should be applicable to a broad population of practitioners and horses.

There are several explanations for variability in this study. One study shows that when horses are only mildly lame, the inter-observer agreement is poor (Keegan *et al.* 1998). Since two thirds of the horses in our population were not markedly positive to flexion, more inter-observer variability is expected. There was no attempt to influence results by including only obvious examples or by excluding practitioner responses. Variability of practitioners' interpretations would also be expected because all practitioners evaluate horses for lameness using more information than just viewing the horses from behind. There is usually a significant time investment in evaluating a horse prior to flexion examination, and in our study, reviewers were only allowed a single view that was only 10-14 seconds in duration. Additionally, our study used a large number of horses which resulted in a large number of videos. It is possible that some clinicians fatigued from watching too many videos in a row. The estimated time to view the series once was approximately 45 minutes. It is also possible that some reviewers did not watch all cases at one sitting, which may add variability to interpretations.

Many lameness studies use video to document lameness exams (Hewetson *et al.* 2006; Keegan *et al.* 1998; Pleasant *et al.* 1997). There are many benefits to using video, and a few limitations. The reviewers were effectively blinded to the duration of flexion which decreased bias. The use of videos allowed us to include practitioners from many different locations and for each of the reviewers to see the exact same thing. Our lameness videos offered only one view of the lameness examination, which was from directly behind the horse. This was chosen because it is the most common position of a practitioner following a flexion test. Additionally, there was no sound, which some clinicians may rely upon when judging lameness.

The results of this study do not include a diagnosis for each patient or what the authors considered the 'true' lameness or 'true' response to flexion tests. The aforementioned lack of a standard prevents us from determining the true response to flexion tests without substituting our own opinion for a standard which is not appropriate because our abilities in no way exceed those of the reviewers. This study was not intended to standardize flexion tests. Our goal was to determine if full hindlimb flexion of five and sixty seconds resulted in the same interpretation; and this does not require a standard. The diagnosis and/or source of pain is also unimportant for the interpretation of the flexion test because in practice, the diagnosis is unknown at the time flexions are being performed.

In conclusion, there is great variability between practitioners as to what constitutes a positive flexion test. We reject our hypothesis that flexion of five seconds will result in the same interpretation as flexion of sixty seconds. While it is true that sixty second flexion was more likely to be judged positive; reviewers' results from five seconds and sixty seconds agreed on average 74% of the time with some reviewers agreeing on the interpretation of flexion for five and sixty seconds in up to 90% of the flexions. Flexion tests of shorter duration may have some utility for some practitioners as a method to abbreviate the examination process; however practitioners should perform their own trial to determine if it is appropriate for their circumstances.

References

Arkell, M., Archer, R.M., Guitian, F.J. and May, S.A. (2006) Evidence of bias affecting the interpretation of the results of local anaesthetic nerve blocks when assessing lameness in horses. *Vet Rec* 159, 346-349.

Beeman, G.M., Soule, S.G. and Swanson, T.D. (1992) History and philosophy of the medical examination of horses for purchase. *Vet Clin North Am Equine Pract* 8, 257-267.

Busschers, E. and van Weeren, P.R. (2001) Use of the flexion test of the distal forelimb in the sound horse: repeatability and effect of age, gender, weight, height and fetlock joint range of motion. *J Vet Med A Physiol Pathol Clin Med* 48, 413-427.

Dawson, N.V., Arkes, H.R., Siciliano, C., Blinkhorn, R., Lakshmanan, M. and Petrelli, M. (1988) Hindsight bias: an impediment to accurate probability estimation in clinicopathologic conferences. *Med Decis Making* 8, 259-264.

Dyson, S. (1997) An approach to hindlimb lameness 2. Gait assessment, flexion tests and what to do next. *In Practice* 19, 14-19.

Dyson, S.J. (1998) Evaluation of the musculoskeletal system. Part 4. The use of flexion tests and small diameter lunging. In: *British Equine Veterinary Association manual: The pre-purchase examination*, Eds: T.S. Mair and B.E.V. Association., Equine Veterinary Journal Ltd., Newmarket, Suffolk, U.K. pp 95-100.

Fuller, C.J., Bladon, B.M., Driver, A.J. and Barr, A.R. (2006) The intra- and inter-assessor reliability of measurement of functional outcome by lameness scoring in horses. *Vet J* 171, 281-286.

Hertwig, K.H. (1850) *Praktisches handbuch der chirurgie für thierärzte*, A. Hirschwald, Berlin,. pp xii, 811, [811] p.

Hewetson, M., Christley, R.M., Hunt, I.D. and Voute, L.C. (2006) Investigations of the reliability of observational gait analysis for the assessment of lameness in horses. *Vet Rec* 158, 852-857.

Kane, A.J. and National Animal Health Monitoring System (U.S.) (2000) *Lameness & laminitis in U.S. horses*, U.S. Dept. of Agriculture, Animal and Plant Health Inspection Service, Veterinary Services, National Animal Health Monitoring System, Fort Collins, CO. p 34 p.

Kaneene, J.B., Ross, W.A. and Miller, R. (1997) The Michigan equine monitoring system. II. Frequencies and impact of selected health problems. *Prev Vet Med* 29, 277-292.

Keegan, K.G. (2007) Evidence-based lameness detection and quantification. *Vet Clin North Am Equine Pract* 23, 403-423.

Keegan, K.G., Dent, E.V., Wilson, D.A., Janicek, J., Kramer, J., Lacarrubba, A., Walsh, D.M., Cassells, M.W., Esther, T.M., Schiltz, P., Frees, K.E., Wilhite, C.L., Clark, J.M., Pollitt, C.C., Shaw, R. and Norris, T. (2010) Repeatability of subjective evaluation of lameness in horses. *Equine Vet J* 42, 92-97.

Keegan, K.G., Wilson, D.A., Wilson, D.J., Smith, B., Gaughan, E.M., Pleasant, R.S., Lillich, J.D., Kramer, J., Howard, R.D., Bacon-Miller, C., Davis, E.G., May, K.A., Cheramie, H.S., Valentino, W.L. and van Harreveld, P.D. (1998) Evaluation of mild lameness in horses trotting on a treadmill by clinicians and interns or residents and correlation of their assessments with kinematic gait analysis. *Am J Vet Res* 59, 1370-1377.

Keg, P.R., van Weeren, P.R., Back, W. and Barneveld, A. (1997a) Influence of the force applied and its period of application on the outcome of the flexion test of the distal forelimb of the horse. *Vet Rec* 141, 463-466.

Keg, P.R., van Weeren, P.R., Schamhardt, H.C. and Barneveld, A. (1997b) Variations in the force applied to flexion tests of the distal limb of horses. *Vet Rec* 141, 435-438.

Lacroix, J.V. (1916) *Lameness of the horse*, American journal of veterinary medicine, Chicago,. pp 8 p.l., 15-261, [210] p.

Liautard, A.F.A. (1888) *Lameness of horses and diseases of the locomotory apparatus*, W. R. Jenkins, New York,. pp 1 p.l., [5]-113 (i.e. 313) p.

Martin, S.W. and Bonnett, B. (1987) Clinical epidemiology. *Can Vet J* 28, 318-325.

Pleasant, R.S., Moll, H.D., Ley, W.B., Lessard, P. and Warnick, L.D. (1997) Intra-articular anesthesia of the distal interphalangeal joint alleviates lameness associated with the navicular bursa in horses. *Vet Surg* 26, 137-140.

Ramey, D.R. (1997) Prospective Evaluation of Forelimb Flexion Tests in Practice: Clinical Response, Radiographic Correlations, and Predictive Value for Future Lameness. In: *Proceedings of the Annual Convention of the AAEP*. pp 116-120.

Ross, M.W. (2003) Manipulation. In: *Diagnosis and management of lameness in the horse*, Eds: M.W. Ross and S.J. Dyson, W. B. Saunders, Philadelphia, Pa. ; London. pp 74-79.

Stashak, T.S. (2002) Examination for Lameness. In: *Adam's lameness in horses*, 5th edn., Ed: T.S. Stashak, Lippincott Williams & Wilkins, Philadelphia. pp 143-147.

Verschooten, F. and Verbeeck, J. (1997) Flexion test of the metacarpophalangeal and interphalangeal joints and flexion angle of the metacarpophalangeal joint in sound horses. *Equine Vet J* 29, 50-54.