Received: 4 October 2020 | Accepted: 9 December 2021

DOI: 10.1111/evj.13545

Equine Veterinary Journal WILEY

ORIGINAL ARTICLE

Visual lameness assessment in comparison to quantitative gait analysis data in horses

Aagje M. Hardeman[1,2] | Agneta Egenvall[3] | Filipe M. Serra Bragança[2] | Jan-Hein Swagemakers[1] | Marc H. W. Koene[1] | Lars Roepstorff[4] | Rene van Weeren[2] | Anna Byström[4]

[1]Tierklinik Luesche GmbH, Lüsche, Germany

[2]Department of Clinical Sciences, Faculty of Veterinary Medicine, Utrecht University, Utrecht, Utrecht, The Netherlands

[3]Department of Clinical Sciences, Swedish University of Agricultural Sciences, Uppsala, Sweden

[4]Department of Anatomy, Physiology and Biochemistry, Swedish University of Agricultural Sciences, Uppsala, Sweden

Correspondence
A. M. Hardeman, Tierklinik Luesche GmbH, Essenerstrasse 39a, 49456 Lüsche, Germany.
Email: aagje.hardeman@gmail.com

Summary

Background: Quantitative gait analysis offers objective information to support clinical decision-making during lameness workups including advantages in terms of documentation, communication, education, and avoidance of expectation bias. Nevertheless, hardly any data exist comparing outcome of subjective scoring with the output of objective gait analysis systems.

Objectives: To investigate between- and within-veterinarian agreement on primary lame limb and lameness grade, and to determine relationships between subjective lameness grade and quantitative data, focusing on differences between (1) veterinarians, (2) live vs video assessment, (3) baseline assessment vs assessment following diagnostic analgesia.

Study design: Clinical observational study.

Methods: Kinematic data were compared to subjective lameness assessment by clinicians with ≥8 years of orthopaedic experience. Subjective assessments and kinematic data for baseline trot-ups and response to 48 diagnostic analgesia interventions in 23 cases were included. Between and within-veterinarian agreement was investigated using Cohen's Kappa (κ). Asymmetry parameters for kinematic data ('forelimb lame pattern', 'hindlimb lame pattern', 'overall symmetry', 'vector sum head', 'pelvic sum') were determined, and used as outcome variables in mixed models; explanatory variables were subjective lameness grade and its interaction with (1) veterinarian, (2) live or video evaluation and (3) baseline or diagnostic analgesia assessment.

Results: Agreement on lame limb between live and video assessment was 'good' between and within veterinarians (median $\kappa = 0.64$ and $\kappa = 0.53$). There was a positive correlation between subjective scoring and measured asymmetry. The relationship between lameness grade and objective asymmetry differed slightly between (1) veterinarians (for all combined parameters, p-values between $P < .001$ and 0.04), (2) between live and video assessments ('forelimb lame pattern', 'overall symmetry', both $P \leq .001$), and (3) between baseline and diagnostic analgesia assessment (all combined parameters, between $P < .001$ and .007).

Main limitations: Limited number of veterinarians (n = 4) and cases (n = 23), only straight-line soft surface data, different number of subjective assessments live vs from video.

Conclusions: Overall, between- and within-veterinarian agreement on lame limb was 'good', whereas agreement on lameness grade was 'acceptable' to 'poor'. Quantitative data and subjective assessments correlated well, with minor though significant differences in the number of millimetres, equivalent to one lameness grade between veterinarians, and between assessment conditions. Differences between baseline assessment vs assessment following diagnostic analgesia suggest that addition of objective data can be beneficial to reduce expectation bias. The small differences between live and video assessments support the use of high-quality videos for documentation, communication, and education, thus, complementing objective gait analysis data.

KEYWORDS
horse, kinematics, motion capture, nerve blocking, orthopaedic examination

1 | INTRODUCTION

The primary goal of most comprehensive lameness examinations is to localise a source of pain in a specific anatomical structure or location. Several different lameness grading scales are currently used, but there is no easily defined and universally accepted system that has good reproducibility and that takes into consideration the substantial spectrum of clinical presentations.[1] Agreement between veterinarians on whether a horse is lame or sound is often low.[2–4] Additionally, following diagnostic analgesia, re-assessment after the intervention may be influenced by expectation bias.[5] Quantitative gait analysis techniques potentially provide highly repeatable and unbiased quantification of lameness and unbiased re-assessment after diagnostic analgesia. Additionally, quantitative gait analysis can serve as reliable documentation and as a basis for communication with clients and colleagues. This is particularly important if the horse is examined on multiple days or by several veterinarians, in legal cases and for insurance claims.

While subjective grading of lameness has been used for a long time, objective assessment is increasingly common. The current study was designed to investigate the relationship between subjective grading and objectively measured vertical movement asymmetries during lameness assessment of straight-line trot. The study focused on assessment of lameness before and after diagnostic analgesia, and on live vs video evaluation, since the latter is increasingly requested by clients and it is also frequently used for research purposes.[3,5–8] We further evaluated whether the relationship between subjective grading and objective measures was consistent throughout a range from sound to moderate lameness.

The main objective of the study was to evaluate the correlation between subjective grading and objectively quantified gait asymmetry and to evaluate if this correlation was consistent between veterinarians, between live and video assessments and between assessments before and after diagnostic analgesia. We hypothesised that there would be a linear correlation between subjective grading and objective asymmetry, but that the regression coefficients would differ (1) between veterinarians, (2) between live vs video assessment, and (3) between baseline assessment and assessment post-diagnostic analgesia interventions. A second objective was to quantify agreement within and between veterinarians to provide a basis for the correlation analysis and to allow for comparison to other studies. Objective gait results may vary depending on which system is being used and that there is a multitude of lameness scales in use around the world. This study provides data using one specific objective gait analysis system in one large referral clinic using a specific protocol for lameness assessment.

2 | MATERIALS AND METHODS

2.1 | Veterinarians and case selection

The study population consisted of 23 cases presented to one equine clinic for lameness evaluation. Four veterinarians were involved, each with ≥8 years of equine orthopaedic experience. Each veterinarian contributed with data from 12 diagnostic analgesia interventions (baseline and post-intervention assessments), performed in one of the 23 cases as part of their normal clinical work. The inclusion criteria for the horses was the presence of a single-limb lameness at trot on a straight line on subjective assessment with a maximum grade of 3 of 5 on an adapted AAEP scale,[9] routinely used at the involved clinic: grade 0 was defined as no lameness, grade 1 as slight lameness at trot only, grade 2 as moderate lameness at trot only, and grade 3 as slight lameness at walk and severe lameness at trot.[9] Half-grades were given if the lameness was perceived to be in between these grades.

2.2 | Protocol

Each of the 23 cases was first assessed live by one of the four veterinarians, followed by a regular lameness workup. During these workups, kinematic data and video footage were recorded simultaneously using a motion capture system and video camera. Data from baseline trot-ups, all diagnostic analgesia interventions subjectively evaluated as positive (defined as lameness improvement of at least 0.5 degrees), and a maximum of one diagnostic analgesia intervention considered negative (no improvement in lameness), were retained for further analysis. Thus, the 12 diagnostic analgesia interventions recorded for each veterinarian consisted of one or more diagnostic analgesia intervention(s) per horse, depending on the course of the lameness workups.

If a horse was subjectively judged to show multi-limb lameness after a diagnostic analgesia, no further data were included from that horse after this evaluation. If there were multiple negative diagnostic analgesia interventions in the same horse, only the last negative intervention was retained for analysis. A flowchart describing the inclusion and exclusion of cases and diagnostic analgesia interventions can be found in Figure 1.

If a diagnostic analgesia intervention was assessed a second time after an additional waiting period (to allow further diffusion of the anaesthetic solution), both assessments were retained. If the examination was continued on a second day, a new baseline assessment was made and all assessments following diagnostic analgesia interventions on this second day were compared to that second baseline. The same protocol was followed if a horse was seen by a different veterinarian on the second day.

A minimum of 3 months from the initial examination, videos were assessed by all participating veterinarians using a 27-inch Dell Ultrasharp screen (resolution 2560*1440, 60 Hz.), individually and blinded to the findings of other veterinarians. Horses were shown in randomised order. For each horse, videos were shown in the same order as during live assessment, ie first the baseline assessment and all diagnostic analgesia assessments in the same order as performed during the live lameness evaluation. Assessors were allowed to watch videos twice but re-evaluation after having proceeded to the next trot-up was not permitted. When the video-assessor did not agree with the live-assessor on the primary lame limb during the baseline trot-up, evaluation of the diagnostic analgesia phases was not performed.

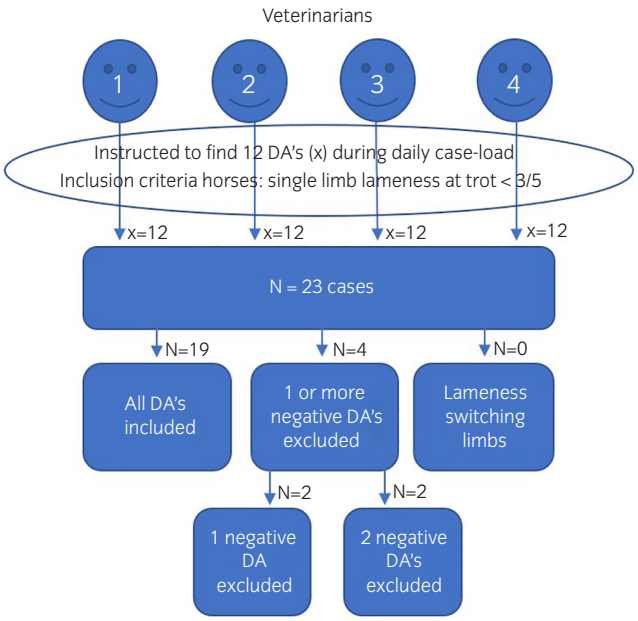

FIGURE 1 Flowchart of the diagnostic analgesia interventions (DA's) per case

During both live and video assessment, lameness observed in each trot-up was graded using the adapted AAEP scale,[9] as detailed above. Assessors were always blinded to the results of the objective measurements, both during live and video assessments.

2.3 | Kinematic data collection

Kinematic data were recorded using Qualisys Motion Capture software (QTM version: 2.14, build: 3180, Qualisys AB), and 20 high-speed infrared cameras (Oqus 700+, 100 Hz sampling frequency, Qualisys AB). The system was set up in a riding arena at the clinic. The total area covered by the cameras was approximately 250 m^2 and the height covered was ≥4 m. Calibration was performed daily according to the manufacturer's protocol. Synchronised video footage (Sony HDR-CX330, 30 Hz) was recorded viewing from the veterinarians' position, ie behind or in front of the horse depending on its direction of travel; using an automated rotating/zooming robot (Pixem & Pixio Robots, Move 'n See). Care was taken to not record veterinarians or owners to facilitate blinding.

The marker setup used was the standard equine clinical marker setup from Qualisys Motion Capture systems (Qualisys AB). Seven spherical reflective markers (25 mm diameter) (Qualisys AB) were placed as follows: a single head marker between the ears, a strip with three markers on the withers (one on the highest point and two at 20 cm lateral to each side of the central one), and a T-shaped strip with three markers located at the tuber sacrale and the craniodorsal aspects of both tubera coxae (Figure 2). Markers were placed by a veterinarian, or a technician experienced with marker placement. Markers remained on the horse between baseline trot-up and the evaluation of diagnostic analgesia but were removed and placed again if the lameness evaluation continued on a second day.

Horses were trotted at their own preferred speed, on a straight line (twice 30 m, turning outside the captured volume). The surface terrain consisted of a combination of sand and synthetic fibre (Figure 2).

2.4 | Kinematic data analysis

All kinematic data were analysed using custom-made Matlab (The MathWorks Inc) scripts. Filtering (4th order Butterworth high-pass

FIGURE 2 Marker placement utilising standardised rubber strips and double-sided adhesive tape for the withers cluster (one marker located on the highest point of the withers and two markers 20 cm lateral to each side of the central marker) and pelvic cluster (three markers located at the tuber sacrale and the craniodorsal aspects of both tubera coxae). The head marker was placed using an elastic textile mask, on which the marker was positioned with Velcro in the midline of the horse. The yellow circles highlight the marker position. The horse is photographed on the surface terrain used during the study

filter, cut-off frequency of 70% of the stride frequency) and stride segmentation were performed as previously described.[10,11] From the vertical displacement of head, withers, and pelvis; single asymmetry parameters were determined to quantify asymmetry at specific anatomical locations: difference between left and right steps in minimum position (MinDiff), maximum position (MaxDiff), and range up between minimum and maximum positions (RUD). These single parameters were used to calculate combined parameters 'forelimb lame pattern',[12] 'hindlimb lame pattern',[12] 'overall symmetry',[13] 'Vector Sum(VS) Head'[14] and 'Pelvic Sum(PS)'.[14] See Table S1 for definitions of the single parameters and formulas of the combined parameters. The parameters 'forelimb lame pattern' and 'VS Head' were used to evaluate forelimb lameness, and 'hindlimb lame pattern' and 'Pelvic Sum' were used to evaluate hindlimb lameness. The parameter 'overall symmetry' was used for all horses. Single parameters were not evaluated on their own in this study.

2.5 | Data analysis

Objective parameters (Table S1) were tabulated as stride-by-stride and as measurement-mean and compared to the subjective data, by matching subjective evaluations with objective data for the corresponding trot-up (one live and up to four video assessments per trot-up). Objective parameters were analysed as absolute values in order to pool left and right limb lame horses. Analyses were partially made including and excluding baseline assessments with disagreement between live and video as a sensitivity analysis. The further statistical analysis consisted of two parts: agreement analysis and analysis of correlations between subjective grading and objective measurements in millimetres. Open source software RStudio (3.3.1, R-Studio) was used, including packages lme4 (1.1-21), lmerTest (3.1-1), emmeans (1.4.3.01) and ggplot2 (3.2.1).

Between- and within-veterinarian agreements were illustrated using Cohen's Kappa (κ) index, including 95% confidence intervals (CI's). Between-veterinarian agreement on lame limb and lameness grade were evaluated comparing video assessments (all possible pairs of veterinarians) and comparing live assessment and each veterinarian's video assessment, including their own cases. Agreement between veterinarians on lameness grade after diagnostic analgesia was calculated for video assessments. Within-veterinarian agreement on lame limb was calculated comparing live and video assessment of their own cases. For the abovementioned agreement calculations on lame limb and baseline lameness grade, only first-day baseline trot-ups were included. All agreements on lameness grade were evaluated using a weighted κ. Agreement was considered poor ($\kappa \leq 0.3$), acceptable ($\kappa = 0.31$-0.5), good ($\kappa = 0.51$-0.8) or excellent ($\kappa > 0.8$).[3]

Correlations between subjective grading and objective measurements in millimetres were analysed using linear mixed models. The limited size of the dataset did not permit that all hypotheses were addressed in one large multivariable model, hence three different models were made. Horse was entered as random effect in all models.

Outcome variables were stride-by-stride data for the combined parameters 'forelimb lame pattern' (data from cases evaluated as forelimb lame), 'hindlimb lame pattern' (data from cases evaluated as hindlimb lame), and 'overall symmetry' (all cases); and also 'VS Head' (forelimb lameness) and 'Pelvic Sum' (hindlimb lameness) for the first model. Normality of residuals was checked by Q-Q-plots; and homoscedasticity was checked by plotting residuals vs fitted values. The limit for statistical significance was set at $P < .05$. Correction for multiple comparisons was not applied.

The first model (model 1), investigated whether the correlation between subjective grading and objective measurements was reasonably linear, while accounting for differences between veterinarians. This was done to verify whether the inclusion of lameness grade as a continuous variable in models 2 and 3 was justified. This first model was made using subjective scorings from video assessments only and excluded data where veterinarians disagreed with the live assessment. Subjective lameness grade was modelled either as a continuous variable or as a categorical variable to compare between these two alternatives. Veterinarian was included as a random effect.

From model 1, least square means for the combined parameters were estimated for each recorded lameness grade (0-3, in steps of 0.5), with lameness grade modelled as a continuous variable and as a categorical variable to allow for appreciation of the linearity of the correlation. The number of trot-ups and strides per lameness grade was checked to detect if caution in interpretation of findings was needed because of small sample sizes.

A second model (model 2) evaluated differences in the subjective-objective correlation between live and video assessments. This

TABLE 1 Between- and within-veterinarian agreement on lame limb, lameness grade and percentage improvement after diagnostic analgesia during video assessment and for live compared to video assessment (Cohen's Kappa (κ) index/weighted κ for lameness grade; median (written in bold)/minimum/maximum). Minimum (min) and maximum (max) number of observations

			Baseline or after	Number of observations		Kappa		
Agreement	Situation	Agreement on	Diagnostic analgesia	min	max	min	median	max
Between veterinarians	Live vs video	Lame limb	Baseline	23	23	0.52	**0.64**	0.81
		Lameness grade	Baseline	23	23	0.21	**0.25**	0.29
Between veterinarians	Video	Lame limb	Baseline	23	23	0.44	**0.58**	0.63
		Lameness grade	Baseline	23	23	0.13	**0.33**	0.37
		Lameness grade	Diagnostic analgesia	28	38	0.1	**0.38**	0.46
Within veterinarians	Live vs video	Lame limb	Baseline	5	7	0.38	**0.53**	1.00

model was made both including and excluding video evaluations where the veterinarian disagreed with the live assessment on the lame limb as a sensitivity analysis. The interaction between lameness grade, entered as a continuous variable, and live or video assessment, entered as a categorical variable, was modelled as fixed effects. Veterinarian and the interaction between veterinarian and lameness grade were included as random effects.

A third model (model 3) evaluated differences in the subjective-objective correlation between baseline assessments and assessments after diagnostic analgesia, and quantified differences in this correlation between veterinarians. This model was made using video assessments, including the evaluations where the veterinarian disagreed with the live assessment on the lame limb. The interaction between lameness grade entered as a continuous variable and baseline or diagnostic analgesia assessment and veterinarian and the interaction between veterinarian and lameness grade were included as fixed effects. Least square means were estimated for each lameness grade and veterinarian.

3 | RESULTS

3.1 | Descriptive data

The 23 cases included were 22 warmbloods >1.60 m, body mass range 480-670 kg (mean 592), age range 5-17 years (mean 10). Fifteen cases were subjectively assessed as forelimb lame (6 left/9 right), and 8 as hindlimb lame (3 left/5 right). Initial lameness grade during live assessment was 1.5 out of 5 (mean, median). Four horses had a lameness evaluation that was performed over two days (3 forelimb lame horses and 1 hindlimb lame horse). Another horse was seen by two different veterinarians on two consecutive days. A total of 48 diagnostic analgesia interventions were included (mean 2.1/median 2 per horse). One diagnostic analgesia was assessed twice with 10 minutes in between. A frequency table of the diagnostic analgesia interventions classified by anatomical location is given in Table S2.

From two horses, one negative diagnostic analgesia was excluded, and from two other horses, two negative diagnostic analgesia interventions were excluded for analysis. No diagnostic analgesia interventions were excluded due to lameness switching to another limb (Figure 1). Seventy-five straight-line live assessments were included, of which 27 were baseline trot-ups and 48 were trot-ups after diagnostic analgesia. Video recordings were evaluated by all four veterinarians, totalling 257 assessments (of 75*4 video assessments possible, 43 potential assessments of videos from 11 different horses were not performed due to non-agreement with the live assessing veterinarian on lame limb). The mean/median number of strides used for the objective analysis was 19/19, and the mean speed (s.d.) was 3.4(0.26) m/s.

3.2 | Agreement between and within veterinarians

Table 1 shows the results for between-veterinarian agreement on lame limb and lameness grade during baseline trot-up (live and video), on lameness grade and within-veterinarian agreement on lame limb (live and video assessment of the same horse). Median agreement on lame limb between veterinarians was slightly higher when comparing each veterinarian during video assessment to live assessment (κ = 0.64) vs based solely on video assessment (κ = 0.58). Median within-veterinarian agreement, ie assessment of their own cases live vs video, was similar (κ = 0.53) but with a larger range (κ = 0.38-1.00). Median agreement between veterinarians on lameness grade was 'poor' when comparing video assessments to the live assessment (κ = 0.25) and 'acceptable' when based solely on video assessments (κ = 0.33 during baseline trot-up, κ = 0.38 after diagnostic analgesia).

Based on linear mixed models of three objective parameters: 'forelimb lame pattern', 'hindlimb lame pattern' and 'overall symmetry', using video assessments only, there were small, though significant ($P < .05$) differences between veterinarians in millimetres of objective asymmetry per subjective lameness grade (Figure 3, Data S1). For example, for the 'overall symmetry' (Figure 3), a lameness grade of 1 out of 5 is estimated to correspond to 38 mm of asymmetry if given by veterinarian 1, but only to 33 mm if given by veterinarian 4. Differences between veterinarians can also be appreciated

from Kappa (κ) values for between-veterinarian agreement on lameness grade during video evaluation (0.33 for baseline assessment, 0.38 for assessment following diagnostic analgesia, Table 1).

3.3 | Subjective lameness grade vs quantitative gait analysis data (model 1)

Based on a numerical comparison between least square means extracted with lameness grade modelled as a continuous variable vs as a categorical variable (Data S2), and visual assessment of Figure 4, it was concluded that the correlation between subjective grading and objective measurements could be approximated as a linear effect for the continued analysis.

Least square means for the five combined parameters estimated independently for each lameness grade (modelled as categorical variable), based on video assessments, are shown in Table 2 and Figure 4. For example, a forelimb lameness subjectively graded as 2 out of 5 corresponds to an estimated value for the 'VS Head' of 58 (95% CI 49-68) mm.

3.4 | Differences between live and video assessment (model 2)

The linear relationship between subjective lameness grade and objective measurements (mm per lameness grade) differed significantly between live and video assessments (Table 3, Data S3). For example, a horse with a one-degree lameness (subjectively) was predicted to show an 'overall asymmetry' of 27.1 plus 10.9 mm equals 38 mm during video assessment. During live assessment, the same horse was predicted to show 27.1 minus 1.2 plus 10.9 plus 1.6 mm equals 38.4 mm. A one-degree difference in lameness grade corresponded to a smaller difference in objective asymmetry (mm) during live assessment compared to video assessment for both the 'forelimb lame pattern' (7 mm less during live assessment, $P < .001$) and 'overall symmetry' (1.6 mm less, $P = .001$). For the 'hindlimb lame pattern', this effect was not significant ($P = .27$). Additionally, for the 'forelimb lame pattern' lameness grade zero corresponded to an additional 6.7 mm objective asymmetry live compared to video assessment (ie the intercepts were different, $P = .003$); there was no significant difference at grade zero for the two other parameters. Exclusion of video assessments with non-agreement on lame limb did not change these conclusions (Data S4).

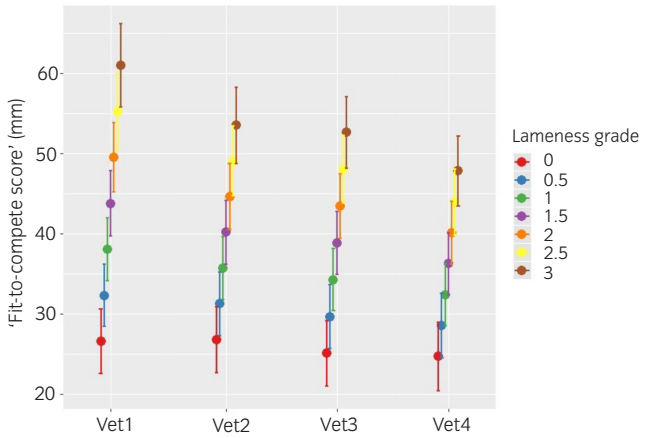

FIGURE 3 Differences between veterinarians based on linear mixed models correlating lameness grade (adapted AAEP scale[9]) to 'overall symmetry' (model 3). Dots indicate least square mean values of the 'overall symmetry' (mm) per veterinarian and grade, vertical error bars indicate 95% CI's. For the 'overall symmetry' formula, see Table S1

3.5 | Differences between baseline assessment and assessment following diagnostic analgesia (model 3)

The linear relationship between subjective lameness grade and objective measurements differed significantly between baseline assessments and assessments after diagnostic analgesia (only video assessments included in the analysis; Table 4, Data S1). For the 'forelimb lame pattern' and 'overall asymmetry', a one-degree difference in lameness grade corresponded to 13.2 mm ($P < .001$) and 1.5 mm ($P = .007$) extra, in terms of objective asymmetry difference, compared to the baseline assessment. In other words, there was a steeper relationship between subjective grade and millimetres measured for assessments after diagnostic analgesia. For the 'hindlimb lame pattern', the opposite was true: a one-degree difference in subjective lameness grade corresponded to 9.7 mm less reduction in the objective asymmetry during assessment following diagnostic analgesia, compared to baseline assessment ($P < .001$) (Table 4).

4 | DISCUSSION

4.1 | Agreement between and within veterinarians

Agreement on lame limb between veterinarians based on live vs video assessments, between veterinarians during video assessment, and within-veterinarian can all be interpreted as 'good'.[3] Differences in assessment conditions (live vs video evaluation and straight line vs lunge evaluation) preclude direct comparison of Kappa values in this study with data from previous studies.[2,3] Nevertheless, there seems to be a better agreement in the current study compared to earlier research. Lower values have been reported on lame limb agreement for video assessment during lungeing (κ = 0.31 between veterinarians, κ = 0.38 for veterinarians with >5 years of experience),[3] and for live assessment on the straight line (κ = 0.37, weighted average experience of 18.7 years).[2] Experience level, frequent teamwork on cases and high-quality videos may have contributed to the higher agreement on lame limb in the current study. This finding contrasts with the agreement on lameness grade, which was classified as acceptable to poor. It should, however, be noted that in Kappa value

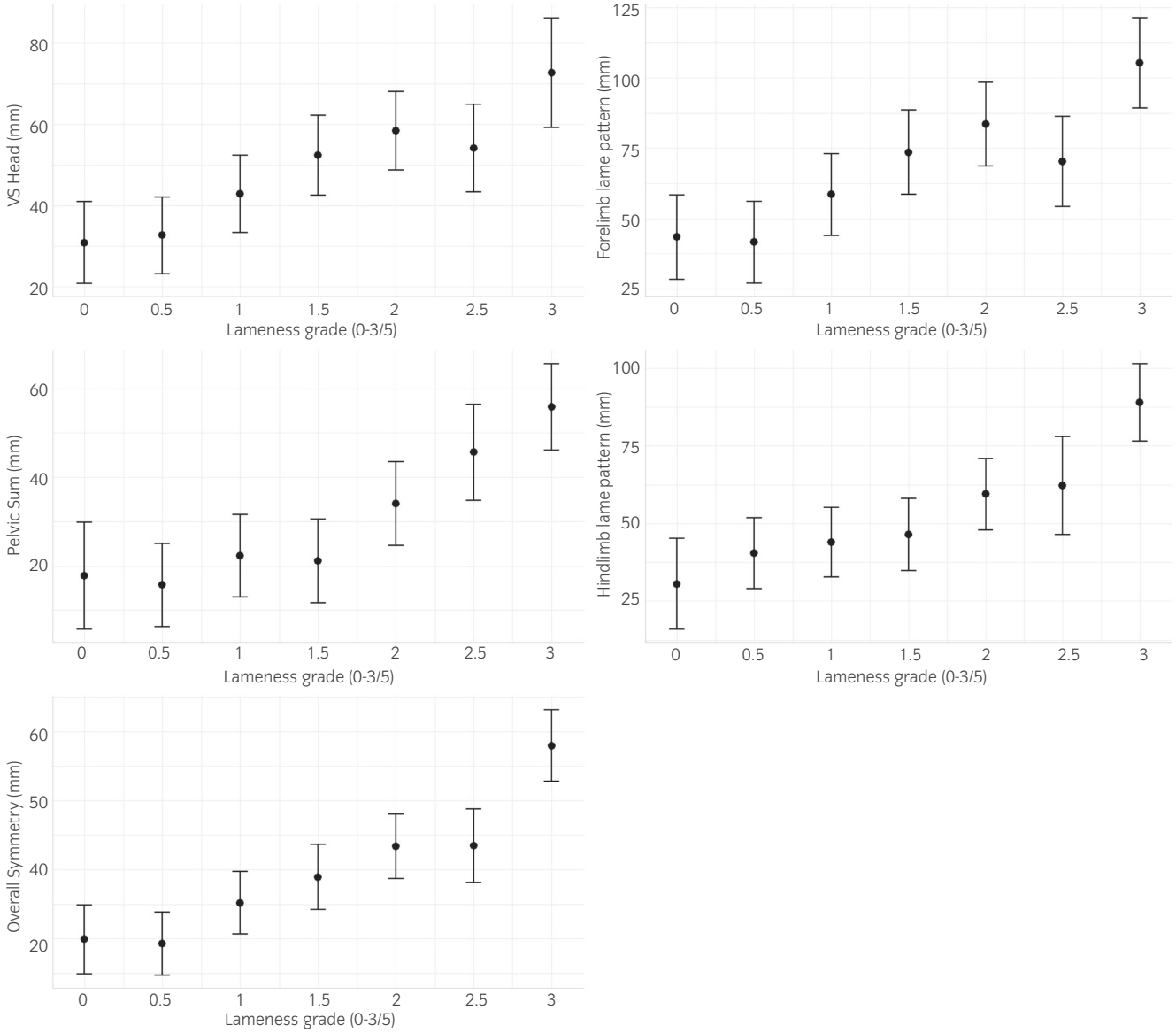

FIGURE 4 Estimated objective asymmetry (least square means with 95% CI's error bars, model 1) for five combined parameters from mixed models correlating subjective lameness grade (adapted AAEP scale[9]) to objective parameters (mm, absolute values, see Table S1 for definitions with formulas)

calculations a disagreement at 0.5 degrees and a disagreement at, for example, 2 degrees, are equally penalised. It can also be questioned whether a higher agreement (with 0.5 degree precision) is achievable for visually grading lameness <3 out of 5, given the known limitations of human visual asymmetry perception.[15]

4.2 | Subjective lameness grade vs quantitative gait analysis data

Irrespective of whether lameness grade was modelled as a categorical variable or as a continuous variable, increments in objective asymmetry between lameness grades were reasonably similar, except for grades 0-0.5 and 2.5 out of 5. For grade 2.5, the results are uncertain due to the low number of occurrences (n = 6 assessments, 88 strides). For very low-grade lameness (0-0.5), scoring is known to be less accurate.[2,15] In addition, the first author (AH) observed that participating veterinarians had difficulties in scoring a 0.5-degree lameness improvement after diagnostic analgesia. These factors may explain why the results for 0.5 deviated from what was expected based on linear prediction. Overall, these results suggest that subjective grading of lameness is relatively linearly related to the horse's movement asymmetry. Hence, the objective asymmetry with a two-degree lameness can be expected to be approximately twice that for a one-degree lameness. However, this conclusion cannot be generalised and needs to be confirmed in larger studies including veterinarians from different clinics and a wider variety of cases. Further, this relationship might not hold true for all scoring systems, given the large differences in lameness workup routines worldwide.[1] The inclusion of objective data could provide a calibration tool for

TABLE 2 Estimated objective asymmetry (mm, least square means and 95% CI's) from mixed models (model 1) with subjective lameness grade (adapted AAEP scale[9]) as explanatory variable, and based on assessments of videos. For the objective parameter definitions see Table S1. Further details see Data S2

Lameness grade	Outcome parameter									
	'VS head'		'Pelvic sum'		'Forelimb lame pattern'		'Hindlimb lame pattern'		'Overall symmetry'	
	mean	CI	mean	CI	mean	CI	mean	CI	mean	CI
0	30.8	20.8-40.8	17.8	4.81-30.8	43.4	28.5-58.4	30.6	15.4-45.8	30.8	26.3-35.3
0.5	32.5	23-42.1	15.7	4.63-26.7	41.6	27.1-56.1	40.4	27.9-52.9	30.4	26.1-34.7
1	42.8	33.3-52.3	22.2	11.27-33.2	58.6	44.1-73.1	44.0	31.6-56.4	35.0	30.7-39.3
1.5	52.3	42.5-62.1	21.1	10-32.2	73.6	58.7-88.5	46.4	33.8-59	38.3	33.9-42.6
2	58.3	48.6-68	34.0	23.01-45.1	83.6	68.7-98.4	59.4	46.8-72	42.2	37.9-46.6
2.5	54.1	43.3-64.9	45.7	33.68-57.7	70.2	54.2-86.3	62.2	45.9-78.4	40.0	35.2-44.8
3	72.7	59.2-86.1	55.9	44.64-67.1	105.3	89.2-121.3	88.9	75.6-102.3	51.9	47.3-56.5

TABLE 3 Results of linear mixed models (model 2) with standard errors (SE), for objective asymmetry (mm) vs the explanatory variables: subjective lameness grade (adapted AAEP scale[9]), and its interaction with live or video assessment

	Outcome parameter					
	'Forelimb lame pattern' (mm [SE])	*P* value	'Hindlimb lame pattern' (mm [SE])	*P* value	'Overall symmetry' (mm [SE])	*P* value
Video						
Lameness grade 0	31.4 (6.4)	<.001	23.0 (5.8)	<.001	27.1 (2.0)	<.001
+1 lameness grade	+36.0 (2.7)	<.001	+26.3 (2.3)	<.001	+10.9 (0.7)	<.001
Difference live – video						
Lameness grade 0	−6.7 (2.2)	.003	−0.1 (3.5)	1	−1.2 (0.7)	.09
+1 lameness grade	+7.0 (1.8)	<.001	+2.5 (2.2)	.3	+1.6 (0.5)	.001

Note: The upper half shows intercepts (lameness grade 0), and regression coefficients (change in objective asymmetry for an increase of +1 lameness grade) for video assessment. The lower half shows differences in intercept and regression coefficient, respectively, for live assessment compared to video assessment. For the definitions of outcome parameters see Table S1. For full model printouts see SupInfo S4

clinicians to achieve better accuracy and consistency in lameness assessment.

4.3 | Differences between live and video assessment

The same reduction in lameness grade corresponded to less reduction in objective asymmetry during live assessment compared to video assessment. We hypothesised that expectation bias[5] is generally larger in the live situation; veterinarians might be more inclined to see the expected or preferred outcome when assessing their own cases. Another possible factor for the discrepancy in grading between live and video assessment is that during video assessment veterinarians were unaware of the history of the case. To minimise risk that veterinarians would recognise their own cases during video evaluation, there was a 3-month period between live and video evaluation and no veterinarians or owners were visible in the video footage. When asked immediately after video assessments, veterinarians responded that they had not recognised specific horses, except for one case by one veterinarian. The potential for the sound produced by an asymmetric gait[16] was unlikely to play a substantial role in the current study, as evaluations were performed on soft surface.

Interestingly, exclusion of video evaluations where veterinarians did not agree with the live assessment on the primary lame limb did not change the conclusions of the statistical models. The fact that the relationship between subjective grade and objective asymmetry did not change despite that a different limb was graded, suggests that veterinarians consider whole body motion, including any compensatory asymmetries. This may lead to a similar grade despite disagreeing on the lame limb and fits well with the widely accepted notion that the human brain is excellent in pattern recognition.[17]

Notwithstanding the fact that there were statistically significant differences between subjective grading and objective measurements comparing between live vs video assessment, the results for live and video assessments were still similar from a practical perspective. For example, a one-grade forelimb lameness corresponded to 36 mm during video assessment vs 29 mm during live assessment (Table 3). These data suggest that high-quality videos can be a valid method for retrospective review of straight-line trot-ups. This is an important conclusion: review of videos allows veterinarians to calibrate their subjective assessment to the objective analysis or to subjective evaluations

TABLE 4 Results of linear mixed models (model 3) with standard errors (SE), for objective asymmetry (mm) vs the explanatory variables: subjective lameness grade (adapted AAEP scale[9]), and its interaction with baseline or diagnostic analgesia assessment

	Outcome parameter					
	'Forelimb lame pattern' (mm [SE])	*P* value	'Hindlimb lame pattern' (mm [SE])	*P* value	'Overall symmetry' (mm [SE])	*P* value
Diagnostic analgesia						
Lameness grade 0	25.7 (6.4)	<.001	29.8 (6.1)	<.001	26.6 (2.1)	<.001
+1 lameness grade	+45.2 (2.9)	<.001	+21.0 (2.9)	<.001	+11.5 (11.5)	<.001
Difference baseline - diagnostic analgesia						
Lameness grade 0	−10.3 (2.6)	<.001	13.3 (3.5)	<.001	−1.7 (0.8)	.02
+1 lameness grade	+13.2 (2.0)	<.001	-9.7 (2.0)	<.001	+1.5 (0.5)	.007

Note: The upper half shows intercepts (lameness grade 0), and regression coefficients (change in objective asymmetry for an increase of +1 lameness grade). The lower part shows differences in intercept and regression coefficient for assessment following diagnostic analgesia compared to baseline assessment. For definitions of outcome parameters, see Table S1. For full model printouts see Data S1

of colleagues, and/or to obtain or provide a second opinion in difficult cases. Further, high-quality videos can be a valuable part of the clinical documentation.[5,18] In education, videos combined with quantitative gait analysis may also be useful for learning to recognise and grade lameness during an orthopaedic examination.

4.4 | Differences between baseline assessment and assessment following diagnostic analgesia

For hindlimb lame horses, a smaller difference in objective asymmetry between lameness grades was found for assessment following diagnostic analgesia than during baseline assessments. This could be due to expectation bias, as reported earlier.[5,18] For the 'forelimb lame pattern' (including data from cases evaluated as forelimb lame), the opposite was true: there was a larger difference in objective asymmetry between lameness grades during assessment following diagnostic analgesia compared to baseline assessment. The reason for this difference between fore- and hindlimb lameness is not directly obvious, but a possible interpretation would be that effectively blocking and/or correct evaluation[4] of (low-grade) hindlimb lameness is more difficult. This finding should be interpreted with caution as only eight hindlimb lame horses were included.

5 | LIMITATIONS

This study involved only four veterinarians, which may not be sufficient to allow generalisation of the results in terms of absolute values in millimetres corresponding to a certain lameness grade. Sample size calculations were not performed. Our adapted AAEP scale contains 0.5 steps which are not well-defined. Only straight-line, soft surface data were used and horses with forelimb lameness were over-represented. The number of evaluations differed between live and video assessment and between lameness grades and was limited for live assessment. The number of measured strides and the number of evaluations were too low to allow for testing the interaction between live vs video and baseline trot-up vs diagnostic analgesia, ie we could not answer the question whether the difference between live and video depending on whether baseline trot-up or trot up following diagnostic analgesia were evaluated. During video baseline assessments, veterinarians knew that a colleague had evaluated the horse as single-limb lame. It was impossible not to disclose this information due to their participation in the live part.

We chose to always compare evaluations following diagnostic analgesia to baseline trot-up. It would have been ideal to compare to baseline trot-up and also to the previous diagnostic analgesia, as the sequential alteration of lameness can provide important information as well. However, a pilot study suggested that asking for this amount of documentation would hamper veterinarians during their clinical workup routine and might affect data quality.

6 | CONCLUSION

Agreement on lame limb was 'good', whereas agreement on lameness grade was 'acceptable' to 'poor'. Quantitative data and subjective assessments were well correlated in a largely linear fashion. The number of millimetres representing one lameness grade differed slightly per veterinarian and per assessment condition. These rather small differences between live and video assessments support the use of high-quality videos for documentation, communication, and education, complementing objective data. Differences between baseline assessment vs assessment following diagnostic analgesia were also small, but significant, suggesting that addition of objective data is potentially beneficial to reduce expectation bias.

ACKNOWLEDGEMENTS

We thank the participating vets of 'Tierklinik Lüsche'; Dr Nadine Blum, Dr Franziska Kremer, and Dr Grigorios Maleas.

CONFLICT OF INTERESTS

No competing interests have been declared.

AUTHOR CONTRIBUTIONS

AM Hardeman contributed to planning of the experiment, data collection, data processing, statistics and preparation of the manuscript. A. Egenvall and A. Byström contributed to data processing, statistics and preparation of the manuscript. FM Serra Bragança and MHW Koene contributed to preparation of the manuscript. JH Swagemakers contributed to data collection and preparation of the manuscript. L. Roepstorff contributed to preparation of the manuscript. PR van Weeren contributed to the planning of the experiment and preparation of the manuscript.

ETHICAL ANIMAL RESEARCH

Research ethics committee oversight not currently required by this journal: procedures were non-invasive.

INFORMED CONSENT

Owners gave consent for their animals' inclusion in the study.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/evj.13545.

ORCID

Aagje M. Hardeman https://orcid.org/0000-0002-0944-2740
Agneta Egenvall https://orcid.org/0000-0002-8677-6066
Filipe M. Serra Bragança https://orcid.org/0000-0001-8514-7949
Lars Roepstorff https://orcid.org/0000-0002-4109-9284
Rene van Weeren https://orcid.org/0000-0002-6654-1817
Anna Byström https://orcid.org/0000-0002-2008-8244

REFERENCES

1. Dyson S. Can lameness be graded reliably? Equine Vet J. 2011;43:379–82.
2. Keegan KG, Dent EV, Wilson DA, Janicek J, Kramer J, Lacarrubba A, et al. Repeatability of subjective evaluation of lameness in horses. Equine Vet. J. 2010;42:92–7.
3. Hammarberg M, Egenvall A, Pfau T, Rhodin M. Rater agreement of visual lameness assessment in horses during lungeing. Equine Vet J. 2016;48:78–82.
4. Starke SD, Oosterlinck M. Reliability of equine visual lameness classification as a function of expertise, lameness severity and rater confidence. Vet Rec. 2019;184:1–8.
5. Arkell M, Archer RM, Guitian FJ, May SA. Evidence of bias affecting the interpretation of the results of local anaesthetic nerve blocks when assessing lameness in horses. Vet Rec. 2006;159:346–9.
6. Brunnekreef JJ, Uden CJTV, Moorsel SV, Kooloos JGM. Reliability of videotaped observational gait analysis in patients with orthopedic impairments. BMC Musculoskelet Disord. 2005;6:1–9.
7. Kawamura CM, de Morais Filho MC, Barreto MM, de Paula Asa SK, Juliano Y, Novo NF. Comparison between visual and three-dimensional gait analysis in patients with spastic diplegic cerebral palsy. Gait Posture. 2007;25:18–24.
8. Starke SD, May SA. Veterinary student competence in equine lameness recognition and assessment: a mixed methods study. Vet Rec. 2017;181:168.
9. Pluim M, Martens A, Vanderperren K, Sarrazin S, Koene M, Luciani A, et al. Short- and long term follow-up of 150 sports horses diagnosed with tendinopathy or desmopathy by ultrasonographic examination and treated with high-power laser therapy. Res Vet Sci. 2018;119:232–8.
10. Serra Bragança FM, Roepstorff C, Rhodin M, Pfau T, van Weeren PR, Roepstorff L. Quantitative lameness assessment in the horse based on upper body movement symmetry: the effect of different filtering techniques on the quantification of motion symmetry. Biomed Signal Process Control. 2020;57:1–12.
11. Roepstorff C, Dittman MT, Arpagaus S, Serra Bragança FM, Hardeman AM, Persson-Sjödin E, et al. Reliable and clinically applicable gait event classification using upper body markers in walking and trotting horses. J Biomech. 2021;114:1–8.
12. Hardeman AM, Egenvall A, Serra Bragança FM, Koene MHW, Swagemakers JH, Roepstorff L, et al. Movement asymmetries in horses presented for pre purchase or lameness examination. Equine Vet J. 2022;54:334–46.
13. Serra Bragança FM, Brommer H, van den Belt AJM, Maree JTM, van Weeren PR, Sloet van Oldruitenborgh-Oosterbaan MM. Subjective and objective evaluations of horses for fit-to-compete or unfit-to-compete judgement. Vet J. 2020;257:1–5.
14. Keegan KG, Wilson DA, Kramer J, Reed SK, Yonezawa Y, Maki H, et al. Comparison of a body-mounted inertial sensor system-based method with subjective evaluation for detection of lameness in horses. AJVR. 2013;74:17–24.
15. Parkes RSV, Weller R, Groth AM, May S, Pfau T. Evidence of the development of 'domain-restricted' expertise in the recognition of asymmetric motion characteristics of hindlimb lameness in the horse. Equine Vet J. 2009;41:112–7.
16. Ross MW, Dyson S. Movement. Diagnosis and management of lameness in the horse. 2nd edn. St. Louis: Saunders Elsevier; 2011. p. 66–78.
17. Haken H, Kelso JAS, Fuchs A, Pandya AS. Dynamic pattern recognition of coordinated biological motion. Neural Network. 1990;3:395–401.
18. Saposnik G, Redelmeier D, Ruff CC, Tobler PN. Cognitive biases associated with medical decisions: a systematic review. BMC Med Inform Decis Mak. 2016;16:1–14.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Hardeman AM, Egenvall A, Serra Bragança FM, Swagemakers J-H, Koene MHW, Roepstorff L, et al. Visual lameness assessment in comparison to quantitative gait analysis data in horses. Equine Vet J. 2022;54:1076–1085. doi:10.1111/evj.13545