

# Agreement between a body-mounted inertial sensors system and subjective observational analysis when evaluating lameness degree and diagnostic analgesia response in horses with forelimb lameness

Porrakote K. Rungsri<sup>1,2</sup>, Wolfgang Staecker<sup>1</sup>, Pitiporn Leelamankong<sup>1</sup>, Roberto J. Estrada<sup>1,3</sup>, Matthias Rettig<sup>1</sup>, Christoph Klaus<sup>1</sup> and C. J. Lischer<sup>1</sup>

<sup>1</sup> Equine Clinic, Faculty of Veterinary Medicine, Free University of Berlin, Berlin, Germany

<sup>2</sup> Department of Companion Animal and Wildlife Clinic, Faculty of Veterinary Medicine, Chiang Mai University, Thailand

<sup>3</sup> Large Animal Hospital, School of Veterinary Medicine, National University, Heredia, Costa Rica

**Summary:** In 24 horses a lameness was assessed on the straight line before and after diagnostic anaesthesia by a body-mounted inertial sensor systems (BMISS) and by two experience veterinarians. For further studies video clips test unit (n = 101) of all the trials were used. The lameness evaluators were blinded from the results of the BMISS. The inter-observers agreement and agreement of lameness evaluation between BMISS and observers were classified into 3 categories; 1) right forelimb lameness or right forelimb lameness greater than left forelimb lameness, 2) left forelimb lameness or left forelimb lameness greater than right forelimb lameness, and 3) sound or equal right and left forelimb lameness. Kappa statistic ( $\kappa$ ), percentage of inter-observers agreement and between BMISS and subjective system (examiners opinion) are reported. The response of anaesthesia agreement was determined by 6 categories between body-mounted inertial sensors system and high experienced observers. This data were analyzed by calculation of the Kendall' tau (b) test. Inter-observer agreement was almost perfect in the clinicians group who assessed the horses in live clinical situation; moderate in the high and in the moderate experienced group and fair in the interns group in video without sound assessment for 85% ( $\kappa = 0.89$ ), 73% ( $\kappa = 0.53$ ), 69% ( $\kappa = 0.44$ ) and 61% ( $\kappa = 0.29$ ) respectively. The detection agreement between objective (BMISS) and subjective assessment was highest for experienced observers in live clinical situation 79% ( $\kappa = 0.58$ ). In video assessment this agreement was moderate for experienced, 74% ( $\kappa = 0.52$ ), and moderate experienced veterinarians 70% ( $\kappa = 0.48$ ) but fair for the group of interns 67% ( $\kappa = 0.40$ ). The average agreement of lameness improvement after anaesthesia between a BMISS and high experienced observers was moderate [ $T_b$  Value = 0.291]. In conclusion, this study supported that the visual lameness detection depends on the experience of the observer. Results of lameness evaluation of horses with a BMISS and subjective lameness evaluation had moderate agreement of lameness detection and had moderate agreement of lameness improvement after anaesthesia even in experienced observers. This system is a practical and useful tool for the objective lameness detection and can help to prove the effects of regional or joint anaesthesia in horses in clinical situation.

**Keywords:** horse / lameness / body-mounted inertial sensors / agreement / anaesthesia response / gait analysis

## Übereinstimmung zwischen einem body-mounted inertial sensors system und subjektiver Beobachter-Analyse während der Beurteilung von Lahmheitsgrad und Ergebnis der diagnostischen Anästhesie bei Pferden mit Vorhandlahmheit

Bei 24 Pferden wurde eine Lahmheit bzw. der Effekt der lokalen Anästhesien auf gerader Strecke objektiv durch ein "body-mounted inertial sensors system" (BMISS) und subjektiv durch 2 erfahrene Pferdetierärzte beurteilt. Für die weiteren Untersuchungen wurden die Videoclips der insgesamt 101 Testsituationen verwendet. Alle Untersucher waren gegenüber den Resultaten des BMISS verblindet. Die Übereinstimmung zwischen den einzelnen Untersuchern sowie die Lahmheitsbeurteilung zwischen dem BMISS und den Beobachtern wurde in drei Kategorien eingeteilt; 1) Lahmheit vorne rechts oder Lahmheit vorne rechts deutlicher als vorne links, 2) Lahmheit vorne links oder Lahmheit vorne links deutlicher als vorne rechts, 3) ohne Lahmheit oder Lahmheit vorne beidseits gleich ausgeprägt. Die prozentuelle Übereinstimmung zwischen den einzelnen Beobachtern sowie zwischen dem Sensor System und den Beobachtern wurde mittels Kappa-Statistik ( $\kappa$ ) analysiert. Der Effekt der lokalen Anästhesie wurde in 6 Kategorien eingeteilt, mittels Kendall'tau (b) Test wurde die objektive (BMISS) und die subjektive Beurteilung durch zwei erfahrene Pferdetierärzten miteinander verglichen. Die Übereinstimmung zwischen den zwei erfahrenen Beobachtern in der klinischen Lahmheitsuntersuchung war ausgezeichnet [85% ( $\kappa = 0.89$ )]. In der Videoanalyse war sie zwischen sehr erfahrenen [73% ( $\kappa = 0.53$ )] und erfahrenen Pferdetierärzten [69% ( $\kappa = 0.44$ )] mäßig und zwischen Anfangsassistenten genügend [61% ( $\kappa = 0.29$ )]. Die Übereinstimmung zwischen der objektiven (BMISS) und der subjektiven Beurteilung war bei den erfahrenen Pferdetierärzten in der klinischen Lahmheitsuntersuchung [79% ( $\kappa = 0.58$ )] am höchsten. In der Videobeurteilung war diese Übereinstimmung bei den sehr erfahrenen [74% ( $\kappa = 0.52$ )] und den erfahrenen [70% ( $\kappa = 0.48$ )] Pferdetierärzten mäßig und bei den Anfangsassistenten genügend [67% ( $\kappa = 0.40$ )]. Die durchschnittliche Übereinstimmung in der Beurteilung nach erfolgter Anästhesie zwischen dem BMISS und sehr erfahrenen Pferdetierärzten war mäßig [ $T_b$ -Wert = 0.291]. Die korrekte Erkennung einer Lahmheit ist abhängig von der Erfahrung des Untersuchers. Die Ergebnisse der Lahmheitsuntersuchung von mit einem BMISS ausgestatteten Pferden waren sowohl in der Übereinstimmung der Lahmheitserkennung zwischen dem BMISS und sehr erfahrenen Pferdetierärzten als auch in der Beurteilung nach diagnostischer Anästhesie mäßig. Das BMISS ist ein praxistaugliches Hilfsmittel zur objektiven Lahmheitserkennung und zur Beurteilung der Veränderung der Lahmheit nach lokalen Anästhesien.

**Schlüsselwörter:** Pferd / Lahmheit / body-mounted inertialsensors / Übereinstimmung / diagnostische Anästhesie / Ganganalyse

**Citation:** Rungsri P. K., Staecker W., Leelamankong P., Estrada R. J., Rettig M., Klaus C., C. J. Lischer (2014) Agreement between a body-mounted inertial sensors system and subjective observational analysis when evaluating lameness degree and diagnostic analgesia response in horses with forelimb lameness. *Pferdeheilkunde* 30, 644-650

**Correspondence:** Porrakote K. Rungsri, Free University Berlin, Equine Clinic, Oertzenweg 19b, 14163 Berlin, p\_kulchaiwat@hotmail.com

## Introduction

Local anaesthesia (blocking) helps to isolate the focus of pain within the limb in horses with lameness (Adam 1974, Schumacher et al. 2004, Schumacher et al. 2013). This requires the veterinarian to assess the amplitude of lameness by observing the horse in motion before and after the nerve block (Adam 1974). There are many reports in the literature that have evaluated blocks using subjective grading or scoring methods (Dyson 1991, Schumacher et al. 2000, Schumacher et al. 2001). However, using a subjective scoring method (Adam 1974, Ross 2003) that categorizes lameness into a limited number of groups such as mild, moderate, or severe can make quantifying small changes difficult, especially in horses with low grade lameness. Un-blinded subjective evaluation of the effectiveness of blocks, which is of necessity the clinical norm, is also susceptible to bias (Arkell et al. 2006). Objective methods of lameness evaluation, because of high sampling rates, have the potential for increased sensitivity of measurement, with more precise differentiation of small differences and without potential bias of the observer (Keegan et al. 1997, Bidwell et al. 2004). Both kinematic (measuring motion) and kinetic (measuring force) techniques have been described and used for objective measurement of equine lameness (Peham et al. 1999, Keegan et al. 1997, Bidwell et al. 2004, Weishaupt et al. 2006), but usually in a setting that is not amendable to regular clinical use. However, recently, several groups around the world have developed body-mounted inertial sensor systems that are relatively easy to use, that provide data immediately, and are sufficiently repeatable and accurate for clinical use (Thomsen et al. 2010, Keegan et al. 2011a, Keegan et al. 2012).

Comparison of these systems to subjective evaluation for detection of lameness in horses has been reported to be only moderate (Keegan et al. 2013), but this has not been studied using “years-of-experience” of the evaluator as a variable. Comparison of these systems to subjective evaluation for detection of change in lameness after blocking has not been reported. The objectives of this study were (1) to estimate the inter-observer agreement on subjective lameness evaluation by veterinarians of different experience levels, (2) to evaluate one body-mounted inertial sensor system for detection of lameness by comparing to subjective evaluation and (3) to subjective assessment of improvement after local anaesthesia by highly-experienced equine veterinarians, respectively.

We hypothesized that agreement between objective inertial sensor and subjective evaluations will be dependent upon years-of-experience and then agreement for improvement of lameness after blocking with experienced veterinarians will be high.

## Materials and methods

### Study design

#### Horses

Twenty-four horses that were presented to the Equine Clinic, Free University Berlin between March and July 2012 for evaluation of forelimb lameness were selected for inclusion in this study. Fourteen displayed left forelimb lameness and 10 displayed right forelimb lameness. Criteria for selection were 1) mild to moderate forelimb lameness (grade 1–3 of subjec-

Score	Description
0	Sound
1	A subtle head nod is observed may be inconsistent at times.
2	Obvious lameness is observed. The head nod are seen consistently
3	Pronounced head nod are seen several centimeters
4	Severe lameness with extreme head nod. The horse can still be trotted
5	The horse does not bear weight on the limb. If trotted, the horse carries the limb. Horses that are non-weight bearing at the walk or while standing should not be trotted.

**Fig. 1** Lameness grading scale adapted from Ross 2003. Scoring from 0 to 5 or sound to non bearing weight are based on observation of the horse at a trot in hand, in a straight line, on a firm or hard surface.

tive score refer to Figure 1) when the horse was evaluated trotting in a straight line on a hard surface, and 2) a positive response (i.e. perceived improvement in lameness) was obtained after either a palmar digital nerve block or subsequent abaxial sesamoid nerve or distal interphalangeal joint blocks. All horses were adult horses with an age range of 4–24 years (mean = 13.7). The group included 12 Warmbloods, 5 Standardbred Trotters, 5 Ponies, 1 Thoroughbred, and 1 Appaloosa Horse. Owner permission for collection of body-mounted inertial data and for its use in this study was obtained for every case. Besides trotting the horse in a straight line for data collection, all horses had complete lameness evaluation. In cases with bilateral forelimb lameness, the limb with the higher lameness degree seen when trotting in the straight line was used in the study. The decision on the worse limb in bilateral cases was made by an experienced clinician blinded to the result of the body-mounted inertial sensor system. Horses with concurrent hind limb lameness, either detected subjectively or measured with the inertial sensor system were excluded from analysis.

### Clinical Evaluation of Lameness

For lameness evaluations each horse was trotted while restrained with a lead rope by the same trained groom in a straight line on a 30m-long concrete surface. Because the inertial sensor system used in this study recommends collecting at least 25 contiguous strides, most horses were trotted back and forth twice on the concrete surface. One experienced clinician, (first observer) who was blinded to the results of the inertial sensor system, additionally completed a clinical examination of the limb including evaluation of hoof and limb conformation, inspection of strength of digital pulse, palpation of the limb, and application of hoof testers. This clinician made the decision about which limb to block. The severity of lameness was then graded by this clinician and a second experienced clinician, blinded to results of inertial sensor evaluation and to each other’s evaluation, using a 0 to 5 scale (Figure 1). These same 2 clinicians graded the response to blocking using a 0 to 6 scale (Figure 2). Each horse was evaluated for lameness by these 2 clinicians and by the inertial sensor system before local anaesthesia (base line lameness) and then 10 to 15 minutes after each nerve block and on the following day baseline lameness and at 2,

Description in words for subjective evaluation (clinical use)	Score	Grading in percentage of improvement of the VS
Negative	1	no improvement to <25 % improvement
Less than 50 % improvement	2	26 – 50 % improvement
More than 50 % improvement	3	51 – 75 % improvement
Positive with residual lameness	4	76 – 99 % improvement
Positive	5	100 % improvement
Positive with switching lameness to the contralateral side	6	100 % improvement

**Fig. 2** Description of degree of improvement of the response to anaesthesia (perineural nerve or DIP joint block) used in clinical situation in Equine Clinic, Faculty of Veterinary Medicine, Free University of Berlin.

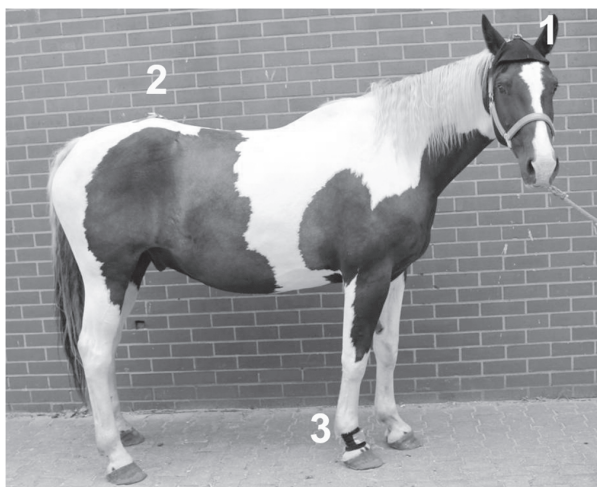
5, and 10 minutes after each joint block. This resulted in 7 trials per horse.

### Diagnostic Anaesthesia

The palmar digital (PD) and abaxial sesamoid (AS) nerve blocks were performed as previously described (Dyson 1984, Schumacher et al. 2004, Schumacher et al. 2013) using 2 ml of 2% mepivacaine (Scandicain®)<sup>b</sup> per site (Schumacher et al. 2004). For the distal interphalangeal (DIP) joint block, the injection site was aseptically prepared and 5 ml of 2% mepivacaine were injected into the joint using the dorsal approach (Schumacher et al. 2004). The PD and AS blocks were considered successful based upon elimination of skin sensation in either the heel bulb region (for the PD block) or around the entire circumference of the coronary band and negative response to hoof tester application across the heels (for the PD block) or around the circumference of the sole (for the AS block). Hoof tester was used to perform all horses before and after blocking.

### Inertial Sensor System and Data Collection

The body-mounted inertial sensor system (BMISS)<sup>a</sup> used in this study works with 3 inertial sensors (Fig. 3). One uni-axial accelerometer, each, was attached to either the head cap in the poll region, or to a piece of hook-and-loop tape fixed on the most dorsal aspect of the pelvis between the tubera sacrale. An additional gyroscopic sensor, which is used only as an event marker to time index right forelimb stance and swing phases, was attached in a special pouch to the dorsal aspect



**Fig. 3** Horse instrumented with the inertial sensor system: 1) Accelerometer at the head, 2) Accelerometer at the pelvic and 3) gyroscope sensor for the right forelimb.

of the right forelimb pastern. The acceleration sensors measure head and pelvic vertical movement asymmetry. This data is collected at 200Hz and transmitted wirelessly to a tablet computer. Data analysis utilizes a fault-detection technique, which is a type of frequency decomposition that saves periodic and discards non-periodic portions of the signal. Periodic components are re-summed and the “fault” identified by detecting vertical head (for forelimb lameness) or pelvic (for hind limb lameness) movement asymmetry. For the purposes of this study only head vertical movement asymmetry was analysed and evaluated.

For measurement of forelimb lameness, vertical head movement asymmetry is reported as the vector sum (VS) (Keegan et al. 2012) of maximum ( $HD_{max}$ ) and minimum ( $HD_{min}$ ) head height difference in millimeters (mm) ( $VS = \sqrt{(HD_{max})^2 + (HD_{min})^2}$ ) between the right and left halves of each forelimb stride. VS represents an overall measurement of head movement asymmetry in both the downward and upward directions, with positive values indicating head movement asymmetry due to right forelimb lameness and negative values indicating asymmetry due to left forelimb lameness. In this 2-component model of head movement evaluation  $HD_{max}$  and  $HD_{min}$  are conceptually related and cannot be separately evaluated. Both the inter-trial repeatability of  $HD_{max}$  and  $HD_{min}$ , which are used to calculate VS, and the 95% confidence interval of the zero y-intercepts versus mean subjective lameness score = 0 (i.e. sound) are approximately 6 mm. Thus trials with  $VS > 8.5$  mm ( $\sqrt{6^2 + 6^2}$ ) are considered “lame” and differences between trials of  $VS > 8.5$  mm are considered “improved” after block (Keegan et al. 2013). VS was assigned either a positive value if the lameness was indicative of a right forelimb lameness (either positive  $HD_{max}$  and  $HD_{min}$ , or negative  $HD_{max}$  and positive  $HD_{min}$ ), or a negative value if the lameness was indicative of a left forelimb lameness (negative  $HD_{max}$  and  $HD_{min}$ , or positive  $HD_{max}$  and negative  $HD_{min}$ ).

Improvement in forelimb lameness was determined by calculating a percent decrease in VS towards baseline using the formula  $(VS_{beforeblock} - VS_{afterblock}) / (VS_{beforeblock} - VS_{threshold}) * 100$ . For comparison to subjective evaluation % improvement was re-classified into the 6 identical intervals defined for subjective evaluation (Figure 3).

### Video Collection and Evaluation

A digital HD videocamera<sup>c</sup> was used to record all straight trotting trials. Video setting and zooming were manually optimized for each horse. The recordings were digitally transferred into a tablet computer and then edited using specialized software. Each video trial was coded with trial number to which the eva-

lators were blinded. One video clip test unit consisted of baseline trial and a corresponding after blocking trial. The horse was trotted up and down on concrete in a straight line before and then after nerve or joint block, which were labelled as such. This resulted in a total of 120 test units from which 101 test units were randomly selected. To help reducing bias towards guessing the forelimb lameness additional 99 video clip test units of horses with hind limb lameness before and after hind limb block were randomly included in the sequence of videos to be reviewed (This data is not shown). A total 200 video clip test units of 20 sessions were evaluated by 13 individual veterinarians. Each video test unit was 3 minutes in long and each evaluator was allowed to see every video sequence twice at real-time speed only. Evaluators looked at the maximum 10 video test unit recordings per sessions. Sound was not included in the video recordings. Each evaluator was blinded to the prior clinical exam results, the inertial sensor results and to each other's results. Video recordings were evaluated by each person in the same room and setting.

The 13 individuals evaluating the video sequences formed 3 distinct groups based on years of experience evaluating lame horses. Four highly experienced (HE) veterinarians with experience ranging from 8–25 years (HE-Group), 4 veterinarians with moderate experience (ME) with 1-3 years of experience (ME-Group), and 5 inexperienced (IE) interns with < 1 year of experience (IE-Group).

*Data analysis*

Differences in each observer's subjective lameness score between right and left forelimbs for each horse and trial were determined by subtracting the score for the left forelimb from the score of the right forelimb. This gave horses with greater right forelimb lameness positive values and horses with greater left forelimb lameness negative values. Subjective scores were thus classified into 3 distinct categories; 1) right forelimb lameness or right forelimb lameness greater than left forelimb lameness, 2) left forelimb lameness or left forelimb lameness greater than right forelimb lameness, and 3) sound or equal right and left forelimb lameness. A total of 144 trials (43 baseline trials and 101 after blocking trials) were used to analyse the Inter-observer agreement and the agreement between subjective and objective evaluated during lameness examination. The 101 test units were use to analysis the agreement for determination of response to anaesthesia.

*Inter-observer agreement*

Inter-observer agreement for the live clinical evaluation (2 evaluators) and for each sub-category of experience for the video evaluation was estimated using Fleiss' Kappa ( $\kappa$ ) stati-

stic (Fleiss 1971, Fleiss and Cohen 1973). Strength of agreement was further described using the benchmark scale of Landis and Koch with  $\kappa > 0$  representing "poor" agreement,  $0.0 < \kappa < 0.20$  representing "slight" agreement,  $0.21 < \kappa < 0.40$  representing "fair" agreement,  $0.41 < \kappa < 0.60$  representing "moderate" agreement,  $0.61 < \kappa < 0.80$  representing "substantial" agreement, and  $0.81 < \kappa < 1.00$  representing "almost perfect" agreement (Landis and Koch 1977).

*Agreement between subjective and objective evaluation*

Inertial sensor evaluation for forelimb lameness evaluation was also categorized into 3 subcategories based upon sign and amplitude of VS. Trials with values of  $VS > 8.5\text{mm}$  were considered right forelimb lameness or right forelimb lameness greater than left forelimb lameness, trials with  $VS < -8.5\text{mm}$  were considered left forelimb lameness or left forelimb lameness greater than right forelimb lameness, and trials with  $-8.5\text{mm} < VS < 8.5\text{mm}$  were considered sound or right forelimb lameness equal to left forelimb lameness. Agreements between subjective evaluation for clinical evaluation and each subcategory of experience with objective inertial sensor evaluation was estimated with Fleiss' Kappa statistic and strength of agreement described with the same benchmark scale described for inter-observer agreement (Landis and Koch 1977).

*Agreement between responses to anaesthesia*

Agreement for determination of response to anaesthesia HE clinical evaluation and inertial sensor evaluation was estimated by calculation of Kendall's Tau-test ( $T_b$ ) rank correlation test. Subjective and objective determinations were considered in agreement only if the classifications matched on improvement score (appendix 2) exactly. Kendall's  $T_b$  ranges from -1, a perfect negative relationship to 1, a perfect positive relationship. Strength of agreement is further described using the benchmark scale of  $|T_b| < 0.1$  indicating "very weak" relationship,  $0.10 < |T_b| < 0.19$  indicating "weak" relationship,  $0.20 < |T_b| < 0.29$  indicating "moderate" relationship, and  $0.30 < |T_b|$  indicating "strong" relationship. Significance of the  $T_b$  statistic for dependence between subjective and objective methods of evaluation was set at  $\alpha = 0.05$  (Agresti 2010, Sen 1968)

**Results**

*Inter-observer agreement*

Inter-observer agreement on forelimb lameness evaluation was generally high, ranging from "almost perfect" ( $\kappa = 0.89$ )

**Table 1** Inter-observers agreement

Mean $\kappa$ value	Clinical situation		Video testing situation	
	Clinicians (n=2)	High-experience (n=4)	Moderate-experience (n=4)	Interns (n=5)
Overall agreement	85 %	73 %	69 %	61%
Expected by chance	0.38	0.38	0.38	0.38
Weighted kappa	0.89	0.53	0.44	0.29
Agreement	Almost perfect	Moderate	Moderate	Fair

for live clinical evaluation to “fair” for video evaluation by interns ( $\kappa = 0.29$ ) (table 1). Live clinical evaluation inter-observer agreement was higher than video evaluation of experts ( $\kappa = 0.53$ , “moderate”) and inter-observer agreement of video evaluation increased with level of experience (Table 1).

*Agreement between subjective and objective evaluations for determination of forelimb lameness*

Percent agreement between subjective and objective evaluation for classification of forelimb lameness for the combined live clinical evaluation of experts and all levels of experience with video review ranged from fair to moderate (Table 2,3). Live clinical evaluation had better agreement with objective evaluation (79% total agreement, mean  $\kappa = 0.58$ , “moderate” agreement) than video review for all experience levels. Agreement between subjective video review and objective measurement increased with years of experience. Evaluators

with high experience (HE) overall had “moderate” agreement with objective evaluation ranging from one individual with 82% total agreement ( $\kappa = 0.61$ , “substantial”) to one individual with 67% total agreement ( $\kappa = 0.41$ , “moderate”). Evaluators with moderate experience (ME) overall also had “moderate” agreement with objective evaluation ranging from one individual with 81% total agreement ( $\kappa = 0.65$ , “substantial agreement”) to another with 64% total agreement ( $\kappa = 0.33$ , “fair agreement”). Inexperienced interns had lowest overall agreement (“fair”) with objective evaluation ranging from one individual with 77% total agreement ( $\kappa = 0.53$ , “moderate”) to one individual with only 45% total agreement ( $\kappa = 0.06$ , “poor agreement”). Overall objective evaluation was less likely to call a horse sound (6.8% of cases) than live clinical evaluation (7.8% of cases), or for video review by highly experienced veterinarians (11.1% of cases), moderately experienced veterinarians (10.8%), or inexperienced interns (11.6%).

**Table 2** The percentage of agreement between the body-mounted inertial sensor measures and subjective lameness score between right and left forelimbs. One hundred forty-four trials were evaluated by 13 observers

Body-mounted inertial sensor measure	Lameness category	Subjective lameness score values			total
		Left > Right(-)	Right > Left(+)	Right=Left(0)	
VS*	Left > Right (-)	76.52%	12.28%	11.20%	100%
	Right > Left (+)	11.87%	73.90%	14.23%	100%
	Right = Left (0)	32.67%	22.00%	45.33%	100%

\*VS = Vector sum

**Table 3** The agreement of lameness detection between a body-mounted inertial sensor system and observers

Mean $\kappa$ value	Clinical situation		Video testing situation	
	Clinicians (n=2)	High-experience (n=4)	Moderate-experience (n=4)	Interns (n=5)
Overall agreement	79 %	74 %	70 %	67 %
Expected by chance	0.42	0.43	0.40	0.39
Weighted $\kappa$	0.58	0.52	0.48	0.40
Agreement	Moderate	Moderate	Moderate	Fair

**Table 4** The percentage of agreement between the body-mounted inertial sensor measures and subjective lameness score of the improvement after blocking of 6 categories (appendix 2) in 101 test units evaluated by 4 observers with high experience (HE)

Improvement after block		HE scores					
		1	2	3	4	5	6
Body-mounted inertial sensor measure score	1	54.29%	18.57%	12.14%	5.71%	7.14%	2.14%
	2	33.87%	30.65%	17.74%	8.06%	9.68%	0.00%
	3	22.73%	27.27%	13.64%	22.73%	13.64%	0.00%
	4	9.09%	27.27%	18.18%	36.36%	4.55%	4.55%
	5*	6.06%	6.06%	12.12%	15.15%	28.79%	31.82%
	6*	-	-	-	-	-	-

\*5 and \*6 of the body-mounted inertial sensor measure score were designed in the same score

**Table 5** Agreement of the response of anaesthesia in 101 test units of 6 categories between the body-mounted inertial sensor measures and the subjective evaluation 4 observers with high experience (HE-Group)

Body-mounted inertial sensors measures	High experience (HE) veterinarians				
	HE1	HE2	HE3	HE4	Average
Tau (t) Value	0.225	0.371	0.185	0.382	0.291
Agreement	Moderate	Strong	Weak	Strong	Moderate

### Agreement between subjective and objective evaluations for determination of improvement of forelimb lameness after blocking

Overall agreement of the HE veterinarians reviewing video with objective evaluation for improvement after blocking was “moderate” ( $T_b = 0.29$ ) but ranged from “strong” for two individuals ( $T_b = 0.38, 0.37$ ) to moderate for one individual ( $T_b = 0.22$ ) to “weak” for one individual ( $T_b = 0.19$ ) (Table 4,5).

## Discussion

The agreement between subjective lameness score of two experienced veterinarians in a clinical situation and the results of an inertial sensor system was moderate, both for detecting the lame limb or the absence of lameness and the assessment of the effect of regional analgesia in horses with mild to moderate forelimb lameness. Even though previous studies have shown a fair to moderate agreement between subjective and objective lameness evaluation (Keegan et al. 2013), our results indicate that there is a moderate agreement between subjective and objective lameness examination, when experienced veterinarians perform live forelimb lameness evaluation, trotting the horses in a straight line on hard surface. The veterinarians used a subjective scoring system to perform the lameness examinations (Figure 1). This same standardized scoring system is used routinely by all the clinicians during the orthopaedic examinations in the Equine Clinic, Berlin. The fact that the 2 evaluators were experienced and used to working together, might have positively affected our results.

We also found that inter-observer agreement for detection of forelimb lameness was higher for live clinical evaluation than for video review and that inter-observer agreement for video review was dependent upon clinical experience. Inter-observer agreement was higher for more experienced veterinarians. Both of these observations are understandable and logical. There are likely many factors that are difficult to completely describe that an experienced veterinarian takes into account when he/she is in the presence of the patient observing and assessing, factors that are mastered by an individual only with experience and that cannot be captured on video. Sound, for example, which was not preserved for video review in this study, of the horse’s hooves hitting the hard surface can be used to detect lameness. This phenomenon has not been studied in detail but it may be both alterations in amplitude and frequency of sound that a “trained” equine veterinarian picks up and utilizes. In another study the intra-observer (Keegan et al. 1998, Fuller et al. 2006), but not the inter-observer agreement was dependent upon experience, suggesting that experience increases consistency (and perhaps confidence) of lameness evaluation, but not necessarily accuracy. However this previous study (Keegan et al. 1998) was limited to horses trotting on a treadmill, which is a highly controlled environment which may not have captured the intangible factors separating experience levels.

Agreement of the inertial sensor system with subjective video evaluation was highest for the “highly experienced” veterinarians and decreased with decreasing level of experience. This is what would be expected if the inertial sensor system was providing relevant evaluation information indicative of fore-

limb lameness. However some individuals in lower experience groups had higher agreement than some individuals in higher experience groups. Agreement with the highly experienced group ( $\kappa = 0.52$ , 74% total agreement) was considered “moderate” but higher than that previously reported using the same inertial sensor system compared to 3 experienced veterinarians performing live lameness evaluation on 106 horses ( $\kappa = 0.41$ , 65% total agreement) (Keegan et al. 2013). The smaller number of horses used in this study along with the other factors previously discussed (working together in the same practice, experience using the same equipment) maybe contributed to this higher agreement.

This is the first study to estimate agreement between this inertial sensor system and experienced veterinarians evaluating forelimb lameness response to blocking. Because the categories for establishing agreement were ordinal with ranking from no response to greater amplitude of response the method of agreement analysis (Kendall’s  $T_b$ ) was different than for simple detection of lameness and determination of side of lameness (Fleiss’ $\kappa$ ). Agreement was moderate for the entire group but substantial for 2 individuals. The results shown the average of  $T_b$  of HE group was moderate. However, two of high experience veterinarians shown significance of  $T_b$  were 0.382 and 0.371, this results indicates the strong agreement of the improvement after blocking of inertial sensor system positively related to the subjective score by HE2 and HE4. So as this inertial sensor system assessment improvement increase, so does the subjective score by HE rating of improve of lameness. On the other hand, the agreement of the improvement after blocking of inertial sensor system and HE veterinarians were going in the same direction as shown you in the table 4. In the horse with score 1 (no improvement to <25% improvement) was clear agreement with the HE score and horse with score 4 (76–99% improvement) and 5 (100% improvement) were shown the same trend. However, score 2 and 3 were difficult to design of HE.

Interestingly, the improvement of the response of anaesthesia compare to the baseline of mild to moderate lameness was difficult to scoring of change after blocking. This study shown the agreement in the 6 categories to classify the response to local anaesthesia in 101 test units between objective and subjective system by the experienced observer group was moderate. However, there was a variation within the experienced observer group as shown in the results. Not only evaluation of mild to moderate lameness detection before but also after administration of local anaesthesia was required to adequately test the accuracy and sensitivity of the objective system. Decreasing the head position or asymmetry of the vertical compare to the baseline was used by veterinarians to identify the response of anaesthesia, which was very important to go further for proximal blocking or to stop investigating the source of pain in clinic situation. This study use the 6 categories to classify the response to local anaesthesia compare to the objective system. If other studies use difference categories/score, it might be shown difference result. This study compare agreement lameness detection evaluated only trotting in a straight line on a hard surface and response of blocking between objective and subjective system. It would be interesting for further study to compare other condition such as in soft or hard surface in circle.

## Conclusions

Results of this study indicated that the detection of mild to moderate lameness and response to regional or joint anaesthesia of horses obtained by use of the body-mounted inertial sensor system-based system did significantly agree with the subjective system, but variation of subjective lameness evaluation based on experience. This study supported that the body-mounted inertial sensors system can be a practical tool for the objective lameness detection and the effects of regional or joint anaesthesia in horses in clinical situation without bias.

## Authors' declaration of interests

No conflicts of interest have been declared. The study is part of the doctoral theses of Porrakote Rungsri, which is still in progress.

## Source of funding

This study was supported by a grant from the Chiang Mai University, Thailand and Free University of Berlin, Germany.

## Acknowledgements

We would like to thank the veterinarians who referred the horses, helped to perform the lameness examination, lameness video test, and supported this study with ideas and suggestions. In special thanks to B. Rheiner, W. Schroeder, S. Jones and all our interns. Moreover, we would like to thank the owners, grooms and horses that make this study possible.

## Authorship

P. K. Rungsri: principal author, study development, data collection, data analysis and manuscript preparation. W. Staeker: study development and data collection. P. Leelamankong: data collection. R. J. Estrada: study development, data collection, manuscript preparation and data analysis. M. Rettig: data, manuscript preparation collection. C. Klaus: study development, data collection, manuscript preparation, C. J. Lischer: senior author, overall study design, project coordination, data analysis and manuscript preparation.

## Manufacturers' addresses

<sup>a</sup> Lameness locator, Equinosis LLC, Columbia, Missouri, USA

<sup>b</sup> Scandicain®2%, AstraZeneca GmbH, Wedel, Germany

<sup>c</sup> digital HD-Videocamera Sony HDV 1080i

<sup>d</sup> Microsoft Windows Live movie maker 2012

## References

Adam O. R. (1974) *Diagnosis of lameness*. 3rd edn., Philadelphia, pp.91-118

Agresti A. (2010) *Analysis of Ordinal Categorical Data*. John Wiley & Sons, pp.419

Arnell M., Archer R. M., Guitian F. J., May S. A. (2006) Evidence of bias affecting the interpretation of the results of local anaesthetic nerve blocks when assessing lameness in horses. *Veterinary Record* 159, 346-348

Bidwell L. A., Brown K. E., Cordier A., Mullineaux D. R., Clayton H. M. (2004) Mepivacaine local anaesthetic duration in equine palmar digital nerve blocks. *Equine veterinary journal* 36, 723-726

Dyson S. J. (1984) Nerve blocks and lameness diagnosis in the horse. *In practice* 6, 102-107

Dyson S. J. (1991) Lameness due to pain associated with the distal interphalangeal joint: 45 cases. *Equine veterinary journal* 23, 128-135

Fleiss J. L. (1971) Measuring nominal scale agreement among many raters. *Psychological Bulletin* 76, 378-382

Fleiss J. L., Cohen J. (1973) The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational Psychologist* 3, 613-619

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

Keegan K. G., Kramer J., Yonezawa Y., Maki H., Pai P. F., Dent E. V., Kellerman T. E., Wilson D. A., Reed S. K. (2011) Assessment of repeatability of a wireless, inertial sensor-based lameness evaluation system for horses. *Am. J. Vet. Res.* 72, 1156-1163

Keegan K. G., MacAllister C. G., Wilson D. A., Gedon C. A., Kramer J., Yonezawa Y., Maki H., Pai P. F. (2012) Comparison of an inertial sensor system with a stationary force plate for evaluation of horses with bilateral forelimb lameness. *Am. J. vet. Res.* 73, 368-374

Keegan K. G., Wilson D. A., Kramer J., Reed S. K., Yonezawa Y., Maki H., Pai P. F., Lopes M. A. F. (2013) Comparison of a body-mounted inertial sensor system-based method with subjective evaluation for detection of lameness in horses. *American journal of veterinary research* 74, 17-24

Keegan K. G., Wilson D. A., Wilson D. J., Smith B., Gaughan E. M., Pleasant R. S. (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

Keegan K. G., Wilson D. J., Wilson D. A., Frankeny R. L., Loch W. E., Smith B. (1997) Effects of anesthesia of the palmar digital nerves on kinematic gait analysis in horses with and without navicular disease. *Am. J. Vet. Res.* 58, 218-223

Landis J. R., Koch G. G. (1977) The Measurement of Observer Agreement for Categorical Data. *Biometrics* 33, 159-174

Peham C., Licka T., Girtler D., Scheidl M. (1999) Supporting forelimb lameness: clinical judgement vs. computerised symmetry measurement. *Equine Vet. J.* 31, 417-421

Ross M. W. (2003) Chapter 7 – Movement. In Mike W. Ross, DVM, Sue J. Dyson M, editor(s). *Diagnosis and Management of Lameness in the Horse*. W.B. Saunders, Saint Louis. pp. 60-73

Schumacher J., Schramme M. C., Schumacher J., DeGraves F. J. (2013) Diagnostic analgesia of the equine digit. *Equine Vet. Educ.* 25, 408-421

Schumacher J., Schumacher J., DeGraves F. J., Steiger R., Schramme M., Smith R., Coker M. (2001) A comparison of the effects of two volumes of local analgesic solution in the distal interphalangeal joint of horses with lameness caused by solar toe or solar heel pain. *Equine Vet. J.* 33, 265-268

Schumacher J., Schumacher J., Schramme M. C., DeGraves F. J., Smith R., Coker M. (2004) Diagnostic analgesia of the equine forefoot. *Equine Vet. Educ.* 16, 159-165

Schumacher J., Steiger R., Schumacher J., DeGraves F. J., Schramme M., Smith R., Coker M. (2000) Effects of analgesia of the distal interphalangeal joint or palmar digital nerves on lameness caused by solar pain in horses. *Vet. Surg.* 29, 54-58

Sen P. K. (1968) Estimates of the regression coefficient based on Kendall's tau. *Am. Statist. Assoc.* 63, 1379-1389

Thomsen M. H., Persson A. B., Jensen A. T., Sørensen H., Andersen P. H. (2010) Agreement between accelerometric symmetry scores and clinical lameness scores during experimentally induced transient distension of the metacarpophalangeal joint in horses. *Equine Vet. J. Supplement* 38, 510-515

Weishaupt M. A., Wiestner T., Peinen K. von, Waldern N., Roepstorff L., Weeren R. van, Meyer H., Johnston C. (2006) Effect of head and neck position on vertical ground reaction forces and interlimb coordination in the dressage horse ridden at walk and trot on a treadmill. *Equine Vet. J. Supplement* 36, 387-392