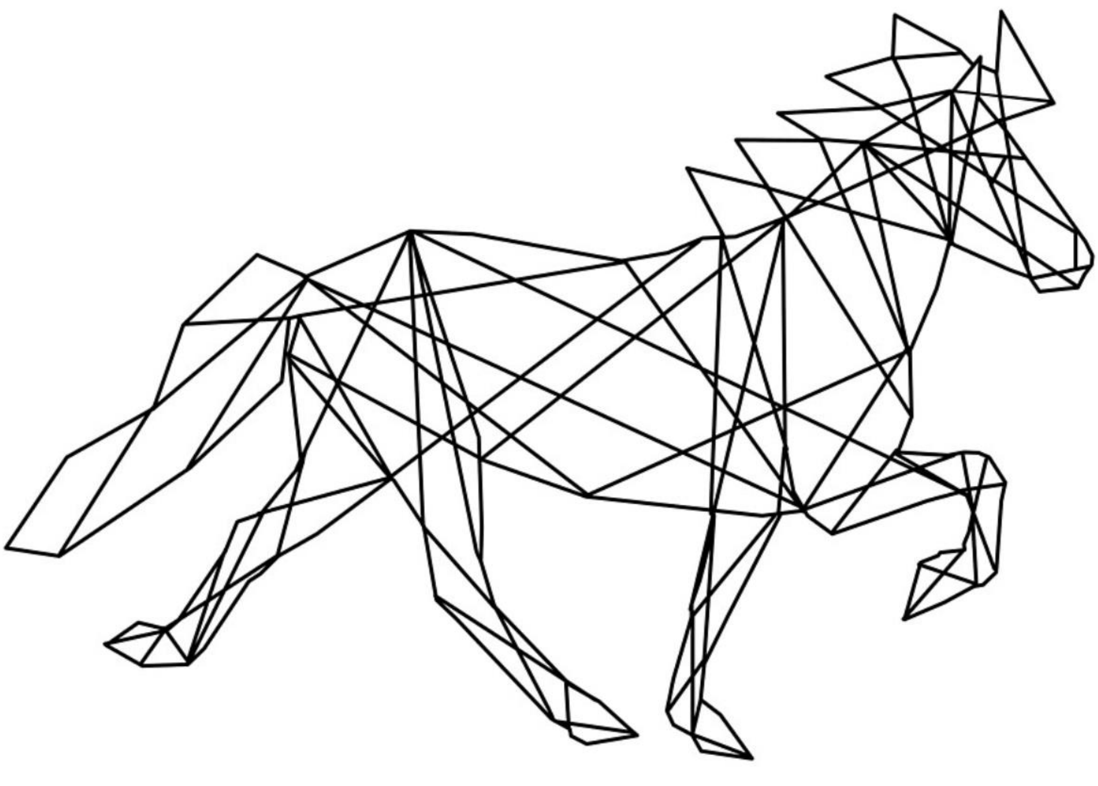
The use of the Equine Utrecht University Scale for Facial Assessment of pain ICELAND (EQUUS-FAP ICELAND) in Icelandic horses at baseline and after mild induced lameness in trot (in hand and ridden) and tölt (in hand)



Master thesis, Department of Equine sciences, Faculty of Veterinary Medicine, Utrecht University, by Jolien Kuis (5635462), written under the supervision of dr. J. P. A. M. (Thijs) van Loon and F.M. (Filipe) Serra Bragança, 2019-2022

Abstract

Objective assessment of pain in horses is a great challenge as animals cannot speak and pain is a personal experience dependent on various factors. Horses being pray animals makes for an even more difficult task, as pray animals often mask pain. Correctly assessing the amount of pain the horse is experiencing is necessary to adequately treat the animal and administer analgesia when needed, leading to more animal welfare and a shorter recovery period. Pain scales are an indispensable tool in assessing equine pain. Pain scales can provide an objective, quantifiable and repeatable way of measuring pain. There are different types of scales, measuring different parameters (e.g. physiological, endocrine and behavioural). One of the types of pain scales validated for several conditions (acute visceral pain, head related pain and pain following orthopedic surgery and trauma) is a pain scale based on facial expressions, the Equine Utrecht University Scale for Facial Assessment of Pain (EQUUS-FAP). This study describes the inter- and intra-observer reliability and the possibility to distinguish the horses at baseline and the horses after receiving lameness induction by total score and separate parameters of an adjusted version of the EQUUS-FAP on Icelandic horses (the EQUUS-FAP ICELAND), ridden and in hand, in trot and tölt, using video footage for scoring. 18 Icelandic horses were used in this study. These horses were filmed in hand (trot and tölt) and ridden (trot) at baseline and after induction of mild lameness. Lameness was induced using a modified horseshoe, placing a M10 screw and applying pressure on the sole of the horse. Four observers scored the footage to determine the inter-observer reliability and two observers were used to determine the intra-observer reliability. The average total EQUUS-FAP ICELAND score of two observers who scored the whole dataset is used to determine the possibility to distinguish the horses at baseline and the horses after receiving induction by EQUUS-FAP ICELAND score. Both the inter- and intra-observer reliability are high (the Intra Class Correlation (ICC) being respectively 0,87 (confidence interval 0,83 – 0,90 p<0,001) and 0,87 (confidence interval 0,80 - 0,91 p<0,001). No significant differences in pain scores between baseline conditions and after induction of lameness were found for both the total average EQUUS-FAP ICELAND score as the separate parameters used (eyelids, nostrils, corners of the mouth/lips/muscle tone, restless mouth and ears). The ridden horses have a significantly higher pain score (p<0,001) than the horses in hand for horses at baseline and for horses after induction in trot. The parameters contributing to this higher score are most likely the nostrils, the corners of the mouth/lips/muscle tone and the restless mouth. There seems to be a rider effect that needs to be further explored.

Table of contents

Introduction

As animals cannot speak, correctly estimating the amount of pain they are experiencing is a difficult task. Pain is a personal experience that is influenced by various factors. The International Association for the Study of Pain (IASP) defines pain as ‘’*An unpleasant sensory and emotional experience associated with, or resembling that associated with, actual or potential tissue damage*.’’ The objective assessment of pain in prey animals can be even more challenging, as they often mask their expressions of pain (Taylor et al., 2002). Furthermore, the subjective nature of the experience leads to an individual and also a breed dependent variation (Wagner, 2010). These factors can possibly lead to the underestimation and suboptimal management of equine pain. This is demonstrated by the fact that even trained veterinarians are often not consistent in assigning pain scores (Price et al., 2002).

The use of a defined pain scoring system can lead to more correct estimation and recognition of pain. This has several benefits. Firstly, recognizing and objectifying pain is necessary to adequately administer analgesia. A scale can help recognize pain early. The early treatment of pain helps to prevent central sensitization and wind-up. This leads to better postoperative analgesia. It is also shown that horses receiving adequate analgesia have a higher food intake post-surgery and recover faster (Taylor et al., 2002; Woolf, 1989). Quantifying a score also helps enhance interobserver agreement, which leads to more consistent evaluations. These more consistent evaluations allow documentation of progress or deterioration (Bufalari et al., 2007). Apart from the physiological benefits, the use of a scoring system also greatly benefits the horse’s welfare by raising awareness of pain. This awareness attributes to the willingness to administer analgesia (Taylor et al., 2002).

Several studies have been conducted, describing the differences in the horses reactions (physiological, endocrine and/or behavioural) to a range of painful conditions (e.g. laminitis, synovitis and colic). These studies have shown that one pain assessment method cannot be sufficient for every situation as the horses experience different kinds of pain (de Grauw & van Loon, 2016). Multiple assessment systems have been developed, and two of them will be further discussed in this section (the grimace scale and the Composite Pain Scale (CPS)). The scale should also be validated for the type of pain, be sensitive enough and, preferably, have a linear relation to the severity of the pain (Ashley et al., 2005).

Pain scales based on facial expressions have been developed and validated for various species (mice and rats) and infant humans (Ahola Kohut et al., 2012; Langford et al., 2010; Sotocinal et al., 2011). These scales have several benefits; Firstly, it is possible to score the horses from a distance: touching the painful area or entering the stable is not necessary, meaning scoring can always be done safely. Secondly, observers can quickly and easily be trained. This opens up the possibility for monitoring outside the clinic. (Dalla Costa et al., 2014).

Composite pain scales consist of several variables (behavioural and/or physiological) that are scored using defined classes (de Grauw & van Loon, 2016). These pain scales have been successfully tested for several kinds of pain (visceral, orthopaedic, inflammatory) and have a high inter-observer reliability (Bussières et al., 2008; Lindegaard et al., 2010; van Loon et al., 2014; van Loon & van Dierendonck, 2015; VanDierendonck & van Loon, 2016). The downsides of these scales are the time consuming nature (especially with repeated measurements), the need for experienced observers and, often, the necessity to manually examen the horses (de Grauw & van Loon, 2016). The CPS also seems to be less effective in the recognition of mild pain, with horses almost never scoring in the high end of the scale, even when experiencing severe pain (Lindegaard et al., 2010; van Loon et al., 2014).

Dalla Costa et al. (2014) developed a grimace scale, comparing its results to a CPS. The participating horses were split in three groups: two groups underwent a routine castration where one group received more extensive analgesia and one group did not undergo the castration. Only the horses that underwent castration showed significantly higher grimace scale and CPS scores. This indicates that the grimace scale can be helpful in the detection of low to moderate pain. Later, the horse grimace scale was also successfully used to asses pain associated with acute laminitis (Dalla Costa et al., 2016). In 2015, Gleerup et al. studied the equine pain face. In this study, pain was induced by placing a tourniquet on the antebrachium and applying capsaicin to the skin. The horses responded to both noxious stimuli with changes in facial expressions. The ears were placed low and/or asymmetrical, the eyes appeared angled, the stare seemed to be tense or withdrawn, there was tension of the chin, lips and facial muscles and the nostrils were dilated mediolaterally.

van Loon & van Dierendonck (2015) developed the Equine Utrecht University Scale for Facial Assessment of Pain (EQUUS-FAP). The horses (control horses and horses with acute colic) were scored using a Visual Analogue Scale (VAS), the EQUUS-FAP and the Equine University Utrecht Scale for Composite Pain Assessment (EQUUS-COMPASS), developed for the composite pain assessment in horses experiencing acute colic pain. Both the EQUUS-FAP and the EQUUS-COMPASS showed a high inter-observer reliability, internal sensitivity and specificity. The EQUUS-FAP was later also validated for acute visceral pain (VanDierendonck & van Loon, 2016), head related pain (van Loon & van Dierendonck, 2017) and pain following orthopedic surgery and trauma (van Loon & van Dierendonck, 2019). A study using the adapted EQUUS-FAP (FOAL) also indicates that a facial expression-based pain scale could be useful in the assessment of acute pain in foals (van Loon et al., 2020). Al mentioned studies in this paragraph were carried out on box rested horses.

Mullard et al., 2017 carried out a study to develop an ethogram that can be used to describe facial expressions in photographs of ridden horses. The goal was also to determine if the ethogram could be used correctly and consistently. The scored features included: the shape and position of the eye, the position of the ears, the position of the mouth, the presence of salivation, the shape of the nostrils, the position of the muzzle, the position of the tongue and the position of the head (tilted, turned). The developed ethogram could be used by observers from different professional backgrounds (veterinarians, veterinary students and horse owners). Following this initial study, (Dyson et al., 2017) used the developed ethogram to determine if it was possible to differentiate between sound and lame ridden horses by facial expression using photographs. It was shown that the pain scores of lames horses were significantly higher than the scores of nonlame horses. The best indicators of pain were: position of the head (severely above the bit, twisting), position of the bit (asymmetrical), ear position (both/one ear backwards) and eye features (exposure of the sclera, (partially) closed, muscle tension, intense stare).

After this study Dyson et al. (2018a) developed an ethogram of a pain scoring system in ridden horses using video footage. Horses experiencing musculoskeletal pain (N=24) and nonlame horses (N=13) were included in this study. The video footage of the horses was assessed by two observers and several of the behavioural markers occurred significantly more frequently in the lame horses. Among these markers are several parameters the EQUUS-FAP uses for scoring (the restless mouth, exposed sclera and the position of the ears). In the same year, Dyson et al (2018b) also carried out a retrospective study on videos of lame horses, before and after analgesia. In this study the total score was significantly reduced after the administering of analgesia. The significant facial parameters included the restless mouth, the exposed sclera and changes in ear position.

As mentioned before, different types of pain (acute vs. chronic pain, somatic vs. visceral pain and nociceptive vs. neuropathic vs. inflammatory pain) cannot optimally be measured using the same scale (de Grauw & van Loon, 2016). Existing scales, like the EQUUS-FAP, should be validated for different noxious stimuli and circumstances (e.g. an horse in the stable, in hand, or ridden). Furthermore the breed of the animal can make a significant difference in the expressing of pain (Wagner, 2010). No previous research was found describing the use of the EQUUS-FAP in ridden horses in general and in Icelandic horses with induced lameness.

The aim of this study is to assess the inter- and intra-observer reliability and the correctness in the prediction of lameness of the altered EQUUS-FAP (the EQUUS-FAP ICELAND) in Icelandic horses with induced lameness. The scale was altered to make it more suitable for the ridden horse. This is necessary as not all of the behaviours normally using the EQUUS-FAP are exhibited while being ridden. As the EQUUS-FAP has proven to be an accurate way of measuring pain in horses with orthopedic pain and ethograms appear to be useful in the indication of lameness in ridden horses, it is important to further investigate the potential of the EQUUS-FAP ICELAND in the detection of lameness in (ridden) horses. The hypotheses of this study are that the EQUUS-FAP ICELAND has a good inter- and intra-observer reliability (ridden and in hand, in trot and tolt) and that the pain scores of the horses after induced lameness will be significantly higher than the pain scores of the horses at baseline (ridden and in hand, in trot in tolt). No significant differences are expected between the different gates (trot and tolt), being in hand or being ridden and the place of induction (hind of front leg).

Materials and methods

Ethical Approval: the study design was approved by the Icelandic Food and Veterinary Authority MAST.

Animals and induction: 18 clinically sound Icelandic horses at Hólar University in Iceland were used for this study. Soundness of the horses was established using EquiMoves, a system for quantitative gait analysis. Among these horses were three mares, nine geldings and four stallions. The horses ranged from riding school horses to sport horses. The horses were colored bay, black and chestnut. Of the 18 horses, two were used as pilot horses, and four were high-performance horses. These six horses (pilot and high-performance were not included in this study. All participating horses were induced in one (n=5) or two different hooves (n=7) (with a minimum of 2 days in between inductions). This induction was established using custom-made shoes with a mediolateral bar (figure 1). By inserting a M10 screw in the hole of this bar and increasing the pressure by tightening it, lameness can be induced. Two experienced veterinarians were used to determine whether the pressure was sufficient. The horses had to be visibly lame, with a score of 2/5 (lameness scale AAEP). The induced lameness is completely reversible.

Figure 1: the modified shoe

Table 1: overview participating horses

Number	**Induction**	**Sex**	**Remarks**
1	Left hind	Gelding	One compromised video
2	None	-	Pilot horse
3	Left front, right hind	Gelding	None
4	Left front	Gelding	None
5	None	Stallion	Performance horse
6	None	Stallion	Performance horse
7	Right front, right hind	Mare	None
8	Left front, left hind	Gelding	No ridden trot
9	Right hind	Mare	None
10	Left hind, right front	Gelding	None
11	Right front, right hind	Mare	None
12	Left front, right hind	Gelding	Four missing videos
13	None	Stallion	Performance horse
14	Right hind	Gelding	None
15	Right front, left hind	Gelding	One compromised video
16	None	-	Pilot horse
17	Right front	Gelding	None
18	None	Stallion	Performance horse

Filming and video footage: Each of the horses is filmed from the side and the back/front simultaneously. In this study, only the footage from the side and the footage of the horses facing the camera was used for scoring. The horses were filmed ridden and in hand. The footage used in this study includes: ridden horses, with and without induction, in fast tölt and in trot; horses in hand, with and without induction, in trot. Not all footage was deemed useable (overview in table 1). The videos were edited to be approximately 10 seconds long, using windows moviemaker. In total, this left 343 videos to score. To assess the intra-observer reliability, 132 films were added twice, adding up to a total of 475 films to score.

The scoring: The footage was randomized and blinded before scoring. The scoring was performed by four scorers with different backgrounds (two trained veterinarians and two veterinary students). Of those four, two observers scored all videos (one trained veterinarian and one student). The other two have scored the videos 1 t/m 40 and 81 t/m 190. Before scoring, a training session of approximately one hour had been performed. This was to ensure the same criteria were used to differentiate between the scores. As one of the scorers lives abroad, a written manual was sent to confirm the same standards were used. The movies were played on a beamer at 20% speed and viewed from a 4 meter distance. After one hour of scoring, a break of fifteen minutes was held. The entire session was never longer than two and a half hours, to ensure concentration and focus of the observers were optimal. As the horses were not filmed in the circumstances the EQUUS-FAP was developed for (rest), some adjustments had to be made to the scoring system (table 1). The head movement; focus; flehming and yawning; and teeth grinding and moaning categories could not be scored. Firstly, some of the behaviour would not be exhibited when riding/walking by hand (e.g. the flehming and yawning). Secondly, because some behaviour would not be visible if displayed under these circumstances (e.g. focus, teeth grinding and moaning). The corners of the mouth and the muscle tone of the head were combined into one category, because it was not possible to see the difference with the bit in. Table 2 (next page) shows the agreements that were made concerning the scoring.

Statistical analysis: Statistical analysis was performed using SPSS and R. Statistical significance was set at P<0,05. As some of the parameters proved to be impossible to score in the movies of the front of the horse, only the videos taken from the side were included in this analysis. In total, this added up to 300 videos, of which 84 were added twice (216 original videos). Intra Class Correlation Analysis (ICC) was used to determine the inter and intra-observer reliability. For the inter-observer reliability, the movies that were added twice were not taken into account (the second scoring was removed). The used footage contains 216 movies scored by two of the assessors, of which 74 were scored by all four of the assessors. For the intra-observer reliability, only the scores of the two assessors who watched all the movies were included. The score of the first scoring was paired with the score of the second scoring, leaving a total of 84 paired values. To test the predictive effect of the pain scale between lame and not lame horses at the different gaits, a linear mixed model was used in R-studio (version 1.3) with the package lme4 (version 1.1) and model estimates (least square means) were calculated using the package emmeans (version 1.4.5). As the intraclass correlation was high, the average scores of the assessors who scored all the movies were used in this model. For each gait, a model was built (trot and tölt separately) with condition (lame and sound), rider (ridden and in hand) and direction (coming back and away from camera) used as fixed effect. Horse ID was used as a random effect. Model adequacy (normality and constancy of variance) was confirmed using visualisation of the scatter plot residuals vs. fitted values and explanatory variables respectively and QQ-plots. Boxplots were made to visualize the effect of the separate parameters on the average scores and the differences between the average scores for horses in baseline, horses after induction, horses in hand and ridden horses.

Table 2: EQUUS-FAP ICELAND

Eyelids	Opened, sclera can be seen in case of eye/head movement	0
	More opened eyes or tightening of the eyelids. An edge of the sclera can been seen 50% of the time	1
	Obviously more opened eyes or obvious tightening of the eyelids. Sclera can been seen >50% of the time	2
Nostrils	Relaxed	0
	Slightly more opened	1
	Obviously more opened, nostril flaring and possibly audible breathing	2
Corners mouth/lips/muscle tone	Relaxed, no fasciculations	0
	Lifted slightly, mild fasciculations	1
	Obviously lifted, obvious fasciculations	2
Restless mouth	Very subtle, almost not visible	0
	One (relatively short) moment of a visible restless mouth	1
	Obviously very restless	2
Ears	Normal, pointed to the front, not behind the neutral line	0
	Behind the neutral line, but <50% of the time	1
	Behind the neutral line >50% of the time	2
Total		…/10

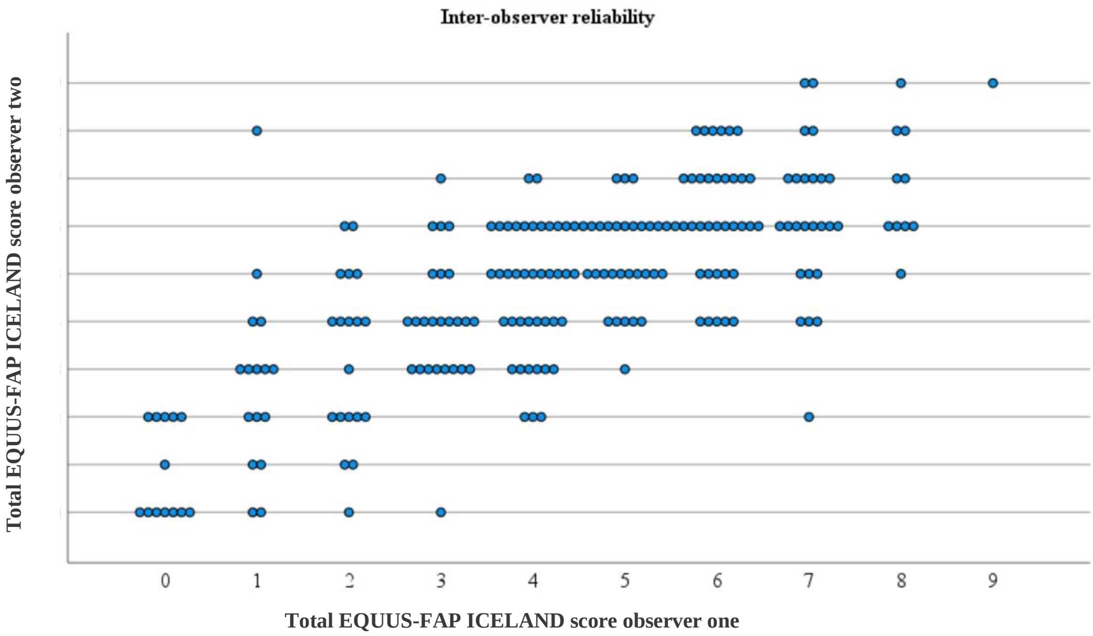
Adjusted, based on: Monitoring acute equine visceral pain with the Equine Utrecht University Scale for Composite Pain Assessment (EQUUS-COMPASS) and the Equine Utrecht University Scale for Facial Assessment of Pain (EQUUS-FAP): A scale-construction study, by: van Loon and van Dierendonck, 2015

Results

Inter-observer reliability

The agreement of the total scores of observers one and two are shown in Figure 2. A significant Intraclass Correlation Coefficient (ICC) is found (p<0,001). The ICC is 0,83 (limits of agreement 0,77 - 0,88). Table 3 contains the ICC's of the individual parameters. Note that the only parameter that does not show a significant correlation is the eyelid (p=0,07). All other parameters show a significant correlation with p<0,001. A part of the videos (n=74) have been scored by 4 observers. The total score of the 4 observers has an ICC of 0,77 (0,48 – 0,89), p=<0,001. ICC values can be split into groups and interpreted as following: a value less than 0,50; a value between 0,50 and 0,75, a value between 0,75 and 0,90 and a value greater than 0,90. Respectively, this categories can be interpreted as: poor, moderate, good and excellent (Koo & Li, 2016). This means that the ICC of the total score in this study indicates a good interobserver reliability.

Figure 2: The agreement of the total EQUUS-FAP ICELAND scores of observer one and observer two



Legend figure 2: Total score Equine Utrecht University Scale for Facial Assessment of pain ICELAND (EQUUS-FAP ICELAND): Intraclass Correlation Coefficient (ICC): 0,83 (0,77 – 0,88), p<0,001 (n=216).

Table 3: The agreement of all the parameters of observer one and observer two

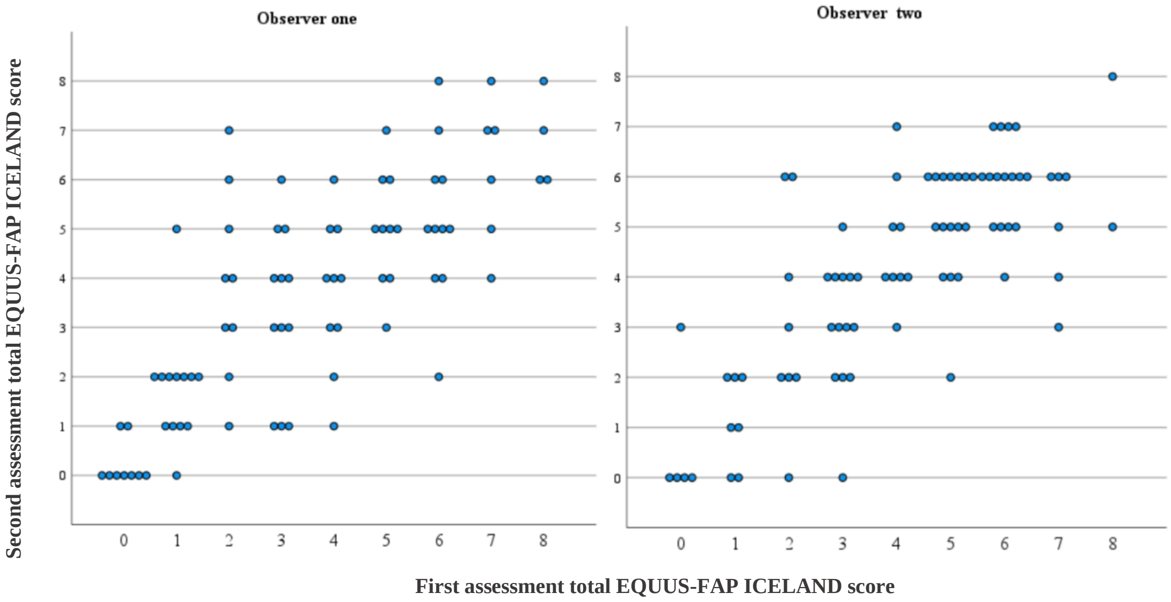
Parameter	**ICC**	**Lower limit of agreement**	**Upper limit of agreement**	***p*-value**
Eyelids	0,21	-0,09	0,43	0,07
Nostrils	0,68	0,56	0,76	**<0,001**
Mouth/lip muscle tone	0,47	0,06	0,67	**<0,001**
Restless mouth	0,84	0,79	0,88	**<0,001**
Ears	0,87	0,83	0,90	**<0,001**
Total	0,83	0,77	0,88	**<0,001**

Legend table 3: Intraclass Correlation Coefficient (ICC), two-way mixed model, average measures (n=216). The significant p-values are sown in bolt.

Intra-observer reliability

The ICC of the total EQUUS-FAP ICELAND scores of the first and the second assessment of the videos is shown in Figure 3 (right: observer one, left: observer two, n=84). There is a significant ICC between both observers. Observer one has an ICC of 0,87 (limits of agreement 0,80 – 0,91) with a *p*-value of <0,001. Observer two has an ICC of 0,87 (limits of agreement 0,80 – 0,92) with a *p*-value of <0,001. Table 4 and Table 5 show the ICC's of the individual parameters. All parameters show a significant correlation (*p*<0,001).

Figure 3: The intra-observer reliability of the average total scores of observer one (left) and observer two (right).



Legend figure 3: Equine Utrecht University Scale for Facial Assessment of pain ICELAND (EQUUS-FAP ICELAND) score first vs second assessment observer one: Intraclass Correlation Coefficient: 0,87 (0,80 – 0,91) (N=84); EQUUS-FAP ICELAND score first vs second assessment observer two: ICC 0,87 (0,80 – 0,92) (n=84)

Table 4: The intra-observer reliability of the EQUUS-FAP ICELAND parameters scores of observer one

Parameter	ICC	Lower limit of agreement	Upper limit of agreement	*p*-value
Eyelids	0,82	0,69	0,90	**<0,001**
Nostrils	0,74	0,56	0,84	**<0,001**
Mouth/lip muscle tone	0,83	0,72	0,90	**<0,001**
Restless mouth	0,85	0,76	0,90	**<0,001**
Ears	0,81	0,71	0,88	**<0,001**
Total	0,87	0,80	0,91	**<0,001**

Legend table 4: Intraclass Correlation Coefficient (ICC) of the Equine Utrecht University Scale for Facial Assessment of pain ICELAND (EQUUS-FAP ICELAND) parameters scores, two way mixed model, average measures (n=84). The significant p-values are shown in bold.

Table 5: The intra-observer reliability of the EQUUS-FAP ICELAND parameters scores of observer two

Parameter	**ICC**	**Lower limit of agreement**	**Upper limit of agreement**	***p*-value**
Eyelids	0,63	0,32	0,80	**<0,001**
Nostrils	0,68	0,47	0,81	**<0,001**
Mouth/lip muscle tone	0,85	0,75	0,91	**<0,001**
Restless mouth	0,90	0,84	0,94	**<0,001**
Ears	0,86	0,78	0,90	**<0,001**
Total	0,87	0,80	0,92	**<0,001**

Legend table 5: Intraclass Correlation Coefficient (ICC) of the : Equine Utrecht University Scale for Facial Assessment of pain ICELAND (EQUUS-FAP ICELAND) parameters scores, two way mixed model, average measures (n=84) The significant p-values are shown in bold.

Induction versus baseline

Model one, trot

A QQ-plot was made, showing normally distributed data as the points fall on the 45-degree reference line (with the exception of one outlier). A Linear Mixed Model is used, calculating estimated marginal means and comparing those values for different conditions (induction (front limb, hind limb), direction of movement (e.g. riding away from the camera or returning) and being ridden or in hand) (Table 6) The Linear Mixed Model (LMM) (n=141) shows a significant difference in the average total score in trot for ridden horses and horses in hand (p<0,001). The ridden horses have higher pain scores than the horses that were not ridden. In the next section, the difference in scores for ridden horses and horses in hand will be further analysed. There is also a significant difference in direction of movement (p=0,01). The induction of both the front and hind limbs show no significant differences in horses in hand (induction front limb p=0,18; induction hind limb p=0,12) and ridden horses (induction front limb p=0,61; induction hind limb p=0,46).

Table 6: ***Results of the LMM model to test the effect of induction (front or hind limb), being ridden or in hand, direction of movement and induction when ridden (front or hind limb) on the total EQUUS-FAP ICELAND score in trot***

Independent	**Predictor**	**Estimate**	**SE**	**t-value**	***p*-value**
Total EQUUS-FAP ICELAND score Trot	(Intercept)	3,17	0,36	8.89	
	Induction front limb	-0,50	0,37	-1,35	0,18
	Induction hind limb	-0,55	0,35	-1,58	0,12
	Ridden	1,94	0,30	6,29	**<0,001**
	Direction	0,55	0,21	2,62	**0,01**
	Induction front limb, ridden	-0,28	0,53	-0,52	0,61
	Induction hind limb, ridden	0,37	0,50	0,73	0,46

Legend table 6: the results of the Linear Mixed Model (LMM) on the effect of induction (front or hind limb), being ridden or in hand, direction of movement and induction when ridden (font or hind limb) on the total Equine Utrecht University Scale for Facial Assessment of pain ICELAND (EQUUS-FAP ICELAND) score. The significant predictors (Ridden and direction) are shown in bold (n=141)

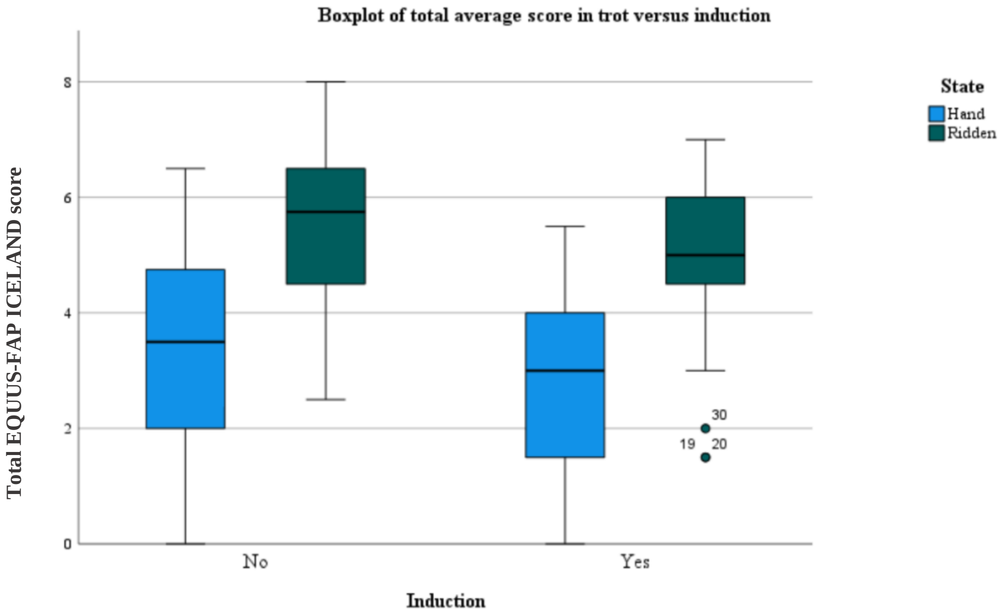
To further analyse the significantly higher scores when ridden, the mean total scores for the EQUUS-FAP ICELAND were compared for: induction (baseline, hind limb or front limb) and being ridden or in hand in trot (table 7). It can be concluded that the ridden horses have a significantly higher EQUUS-FAP ICELAND score in baseline (baseline hand vs baseline ridden: $p<0{,}001$), but also after induction (front limb hand vs front limb ridden: $p<0{,}001$, hind limb hand vs hind limb ridden: $p<0{,}001$). It can also be concluded that ridden horses in baseline do not have a significantly lower score than ridden induced horses (baseline ridden vs front limb ridden: *p*: 0,09, baseline ridden vs hind limb ridden: $p=0{,}70$). The same can been seen for the horses in hand (baseline hand vs front limb hand: $p=0{,}23$, baseline hand vs hind limb hand: $p=0{,}17$). We can also conclude that there is no significant difference in hindlimb and front limb induction, both ridden and in hand (front limb hand vs hind limb hand: $p=0{,}90$, front limb ridden vs hind limb ridden: $p=0{,}23$). A boxplot of the average total scores was made to visualise the difference (Figure 4).

Table 7: ***Results of the Linear Mixed Model to compare the mean total EQUUS-FAP ICELAND score for induction (baseline, hind limb or front limb) and being ridden or in hand, in trot***

		Baseline hand	Front limb hand	Hind limb hand	Baseline ridden	Front limb ridden	Hind limb ridden
Baseline hand	**Estimate**		0,50	0,55	-1,94	-1,17	-1,76
	SE		0,38	0,36	0,32	0,39	0,37
	T ratio		1,31	-6,13	-6,13	-3,005	-4,73
	p-value		0,23	0,17	**<0,001**	**0,005**	**<0,001**
Front limb hand	**Estimate**			0,05	-2,44	-1,67	-2,26
	SE			0,44	0,40	0,45	0,45
	T ratio			0,12	-6,13	-3,70	-4,98
	p-value			0,90	**<0,001**	**<0,001**	**<0,001**
Hind limb hand	**Estimate**				-2,49	-1,72	-2,31
	SE				0,381	0,45	0,42
	T ratio				-6,54	-3,83	-5,52
	p-value				**<0,001**	**<0,001**	**<0,001**
Baseline ridden	**Estimate**					0,77	0,178
	SE					0,40	0,39
	T ratio					1,92	0,45
	p-value					0,09	0,70
Front limb ridden	**Estimate**						-0,59
	SE						0,46
	T ratio						-1,3
	p-value						0,23

Legend table 7: *Results of the Linear Mixed Model to compare the mean total Equine Utrecht University Scale for Facial Assessment of pain ICELAND (EQUUS-FAP ICELAND) score for induction (baseline, hind limb or front limb) and being ridden or in hand, in trot The significant predictors are shown in bold (degrees of freedom range from 136 to 140).*

Figure 4: Boxplot of the total average scores in trot, baseline versus induction, for horses in hand and ridden horses



Legend figure 4: visualization of the significant difference in total Equine Utrecht University Scale for Facial Assessment of pain ICELAND (EQUUS-FAP ICELAND) score for the ridden horses and the horses in hand (N=141). The significant differences are present in baseline (baseline hand vs baseline ridden: p*<0,001) and after induction (front limb hand vs front limb ridden:* p*<0,001, hind limb hand vs hind limb ridden:* p*<0,001).*

Model two, tölt

A QQ-plot was made, showing normally distributed data as the points fall on the 45-degree reference line. A Linear Mixed Model (LMM) is used, calculating estimated marginal means and comparing those values for different conditions (induction (front limb, hind limb) and direction of movement (e.g. riding away from the camera or returning) (Table 8) None of the parameters show a significant result, although the induction of the hind limb has a p value of 0,05. This could suggest a significant result. The horses receiving hind limb induction had a lower total EQUUS-FAP ICELAND score. The other parameters; the front limb induction (p=0,20) and the direction of movement (p=0,73), showed no significant differences. As the horses were only scored in hand, no data is available comparing ridden horses with horses in hand in tölt.

Table 8: ***Results of the LMM model to test the effect of induction (front or hind limb) and direction of movement on the total EQUUS-FAP ICELAND score in tolt for ridden horses***

Independent	Predictor	Estimate	SE	t-value	*p*-value
Total EQUUS-FAP ICELAND score Tolt	(Intercept)	5,51	0,40	13,79	
	Induction front limb	0,50	0,38	1,30	0,20
	Induction hind limb	-0,76	0,38	-2,00	0,05
	Direction	-0,11	0,30	-0,35	0,73

Legend table 8: the results of the Linear Mixed Model (LMM) on the effect of induction (front or hind limb) and direction of movement on the total Equine Utrecht University Scale for Facial Assessment of pain ICELAND (EQUUS-FAP ICELAND) score. There are no significant predictors (n=75)

To further analyse the suspected significant difference in total EQUUS-FAP ICELAND score for the horses with hind limb induction, the mean total scores for the EQUUS-FAP ICELAND were compared for induction (baseline, hind limb or front limb) (Table 9). It can be concluded that the horses with front limb induction have a significantly higher EQUUS-FAP ICELAND score than the horses receiving hind limb induction (*p*=0,02). There is no significant difference in baseline and induced horses (baseline vs front limb induction: *p*=0,21, baseline vs hind limb induction: *p*=0,08)

Table 9: ***Results of the LMM model to compare the mean total EQUUS-FAP ICELAND score for induction (baseline, hind limb or front limb) in tölt***

		Front limb	Hind limb
Baseline	**Estimate**	-0,50	0,76
	SE	0,39	0,39
	T ratio	-1,26	1,95
	p-value	0,21	0,08
Front limb	**Estimate**		1,25
	SE		0,46
	T ratio		2,70
	p-value		**0,02**

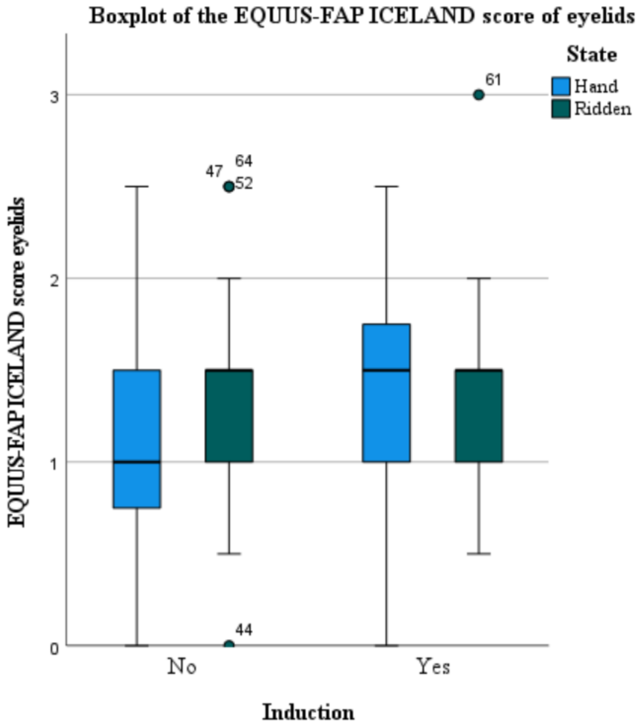
Legend table 9: *Results of the LMM model to compare the mean total total Equine Utrecht University Scale for Facial Assessment of pain ICELAND (EQUUS-FAP ICELAND) score for induction (baseline, hind limb or front limb) in tolt. The significant predictors are shown in bold (degrees of freedom range from 68 to 71)*

Analysis of parameters

To visualise the average EQUUS-FAP scores for the separate parameters, boxplots are made (figures 5 to 9). For these boxplots, only horses in trot were included. This way, it was possible to separate the horses in hand and the ridden horses. Even though no exact conclusions can be made without further statistic analysis, these boxplots give a strong indication that the induced horses do not score significantly higher than the baseline horses on any of the parameters, in hand and ridden.

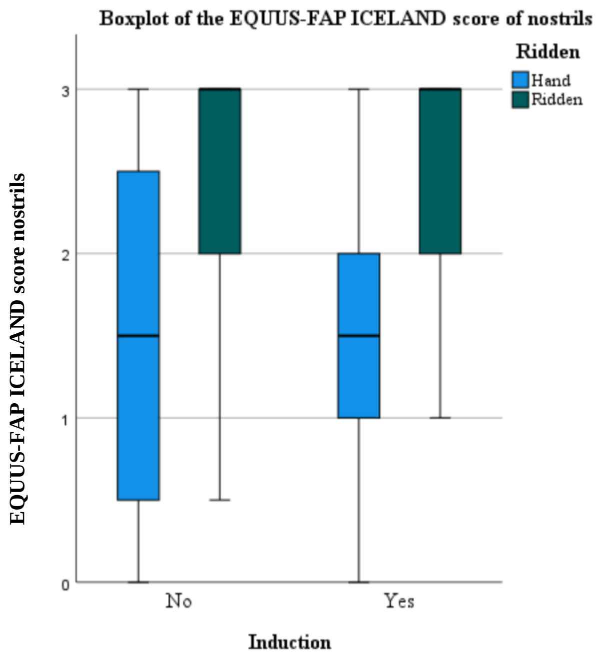
Comparing the ridden horses and horses in hand, the parameters nostrils, mouth/lips/muscle and restless mouth appear to be higher in ridden horses. This suggests these parameters are the biggest contributors to the significantly higher total EQUUS-FAP ICELAND scores in ridden horses.

Figure 5: Boxplot of the EQUUS-FAP ICELAND scores for eyelids in trot, baseline versus induction, for horses in hand and ridden horses



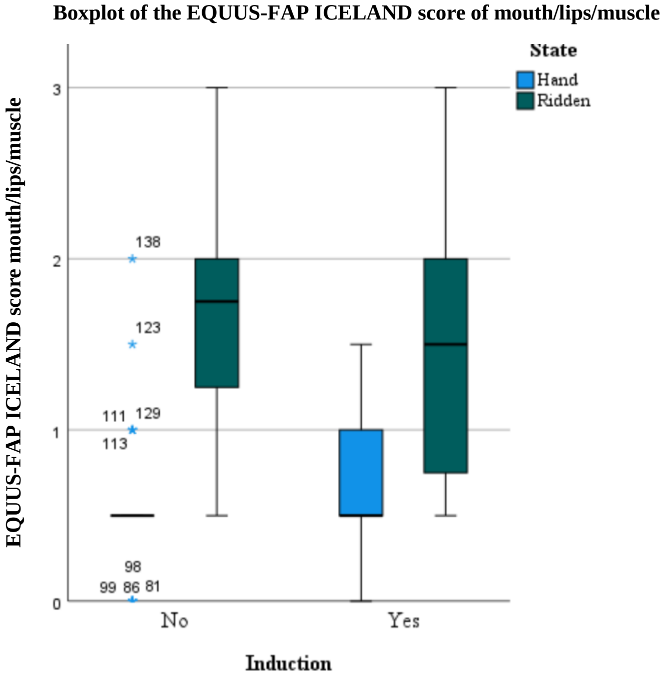
Legend figure 5: visualization of the difference in baseline and induction for Equine Utrecht University Scale for Facial Assessment of pain ICELAND (EQUUS-FAP ICELAND) eyelids score for the ridden horses and the horses in hand (N=96). There appear to be no significant differences between the baseline and the induced horses.

Figure 6: Boxplot of the EQUUS-FAP ICELAND scores for nostrils in trot, baseline versus induction, for horses in hand and ridden horses



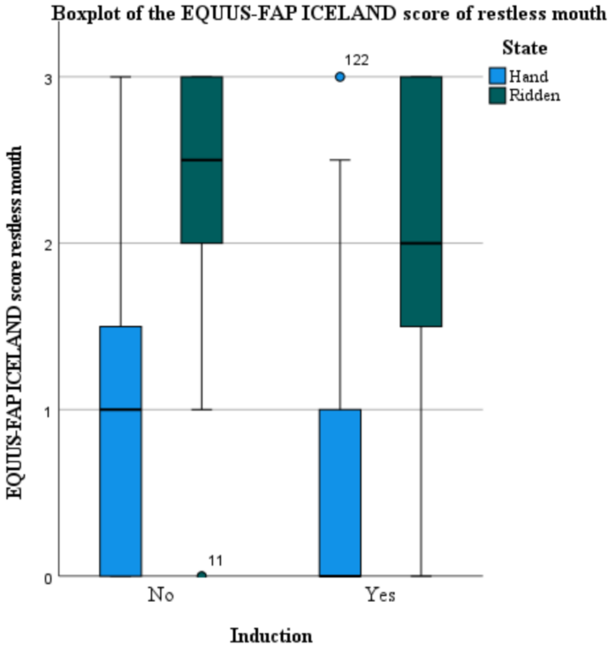
Legend figure 6: visualization of the difference in baseline and induction for Equine Utrecht University Scale for Facial Assessment of pain ICELAND (EQUUS-FAP ICELAND) nostrils score for the ridden horses and the horses in hand (N=122). Note the apparent difference in score for the ridden horses and the horses in hand. There appear to be no significant differences between the baseline and the induced horses.

Figure 7: Boxplot of the EQUUS-FAP ICELAND scores for mouth/lips/muscle in trot, baseline versus induction, for horses in hand and ridden horses



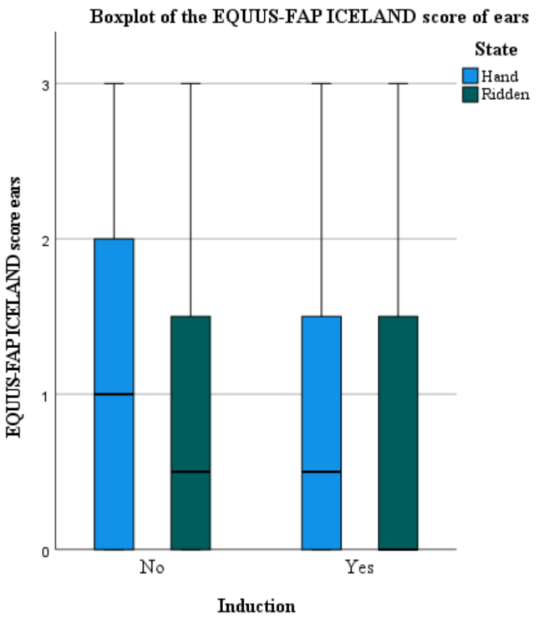
Legend figure 7: visualization of the difference in baseline and induction for Equine Utrecht University Scale for Facial Assessment of pain ICELAND (EQUUS-FAP ICELAND) mouth/lips/muscle score for the ridden horses and the horses in hand (N=114). Note the apparent difference in score for the ridden horses and the horses in hand. There appear to be no significant differences between the baseline and the induced horses.

Figure 8: Boxplot of the EQUUS-FAP ICELAND scores for restless mouth in trot, baseline versus induction, for horses in hand and ridden horses



Legend figure 8: visualization of the difference in baseline and induction for Equine Utrecht University Scale for Facial Assessment of pain ICELAND (EQUUS-FAP ICELAND) restless mouth score for the ridden horses and the horses in hand (N=113). Note the apparent difference in score for the ridden horses and the horses in hand. There appear to be no significant differences between the baseline and the induced horses.

Figure 9: Boxplot of the EQUUS-FAP ICELAND scores for ears in trot, baseline versus induction, for horses in hand and ridden horses



Legend figure 9: visualization of the difference in baseline and induction for Equine Utrecht University Scale for Facial Assessment of pain ICELAND (EQUUS-FAP ICELAND) ears score for the ridden horses and the horses in hand (N=142). There appear to be no significant differences between the baseline and the induced horses.

Discussion

In the current study an altered version of the Equine Utrecht University Scale for Facial Assessment (EQUUS-FAP), the EQUUS-FAP ICELAND, has been used to asses video footage of Icelandic horses in baseline and after induction of lameness in trot (ridden and in hand) and tölt (in hand). Both the inter- and intra-observer reliability are high. However, there no significant difference in total score of the induced horses and the horses at baseline is found (ridden and in hand). The ridden horses do have a significantly higher pain score than the horses in hand at baseline and after induction.

Observer one and two (the observers who scored all the videos) were able to score the video footage similarly, resulting in an significant correlation with an Intra Class Correlation (ICC) (total score) of 0,87. This means that the ICC of the total score in this study indicates a good interobserver reliability. Earlier studies (van Loon & van Dierendonck, 2015, 2017, 2019) validated the EQUUS-FAP for visceral pain, orthopaedic pain and head related pain, achieving high inter-observer reliability in all three studies. It has to be noted that van Loon & van Dierendonck (2015) used video footage and van Loon & van Dierendonck (2017, 2019) used direct observations. It is demonstrated by pilot studies that using video recordings for pain scoring results in lower interobserver reliability and less detailed observations (van Loon & van Dierendonck, 2017). As video footage was used for scoring in this study, it is possible that direct observations would have resulted in an even higher interobserver reliability.

Observer one and observer two (the only observers that scored all the videos), are one trained veterinarian and one veterinary student. A high ICC suggests that the adjusted EQUUS-FAP can also adequately be used by unexperienced individuals. The other two observers also were one veterinary student and one trained veterinarian. The ICC of the videos that were scored by all four assessors, is also high (ICC = 0,77 (confidence interval 0,48-0,89) p<0,001), but slightly lower than the ICC scores of the two assessors. This difference could be observed due to the following reasons: The number of videos scored by all assessors is lower (n=74) than the number of videos scored by two assessors (n=216), resulting in less reliable results. Furthermore, one of the observers is living in Sweden. Therefore, it was not possible to attend the one hour practice session the other observers attended. Although a written manual was sent specifying the conditions used, it is possible that the set-up used for scoring was slightly different (e.g. the quality of the projector, the size of the screen, the length of the scoring session).

Of the 5 parameters used, 4 resulted in a significant correlation (nostrils (ICC = 0,68), mouth/lip muscle tone (ICC = 0,47), restless mouth (ICC = 0,84), ears (ICC = 0,87)), with p values < 0,001. The remaining parameter (eyelids) does not have a significant ICC (0,212, p = 0,07). The eyelids of the horses proved to be difficult to score similarly. One explanation for this could be the colour of the horses. During scoring, it was noticeable that black horses were more difficult to score. The videos used in this study are of the complete horse (so there was no zoom on the head). It was difficult to distinguish the eyelids from the black skin. It would be interesting to further look into the colours of the horses and the agreement of the eyelid scores. If the black horses score significantly more inconsistent, the parameter could still be clinically useable in horses with lighter colours or environments with better lightning or direct observations.

The intra-observer reliability of the total scores is also high in both observer one and two (respectively ICC = 0,87 (confidence interval 0,80 – 0,91), p<0,01 and ICC = 0,87 (confidence interval 0,80 – 0,92), p<0,01). Furthermore, all parameters (eyelids, nostrils, mouth/lip muscle tone, restless mouth and ears) have a high ICC and a p-value < 0,001. This means that both observers were able to consequently score the videos added twice for each individual parameter. Previous studies carried out by van Loon & van Dierendonck (2015, 2017, 2019) did not include an intra-observer reliability because the horses were scored from live observation. However, the intra-observer reliability is calculated for the adapted version of the EQUUS-FAP, made suitable for foals (van Loon et al., 2020). This study also uses video footage for scoring and has a high intra-observer reliability.

Because of the similar looking horses, the high amount of videos and the randomisation, it was impossible to recognize the video or remember the first score that was given. The videos that were scored by the other two observers were not included in the intra-observer reliability, as their number was small. Because of the random order of the videos, many videos that were added twice, were only scored once by the other two observers. Although this high intra-observer reliability shows promise, the results would be more reliable if we used more observers than two and those observers scored the whole set of videos.

Contrary to the hypothesis, the pain scores of the horses after induction of lameness are not significantly higher than the scores at baseline. This is the case for both ridden horses and horses in hand. The separate parameters (eyes, nostrils, mouth/lips/muscle tone, restless mouth and ears) also appear to have no difference for the horses at baseline and after induction. In previous studies, the EQUUS-FAP scores of horses experiencing pain have been significantly higher than the scores of the control horses (van Loon et al., 2020; van Loon & van Dierendonck, 2015, 2017, 2019; VanDierendonck & van Loon, 2016). Therefore it was expected that the EQUUS-FAP scores would be able to adequately distinguish the horses before and after induction of lameness. This difference can be explained by several factors that differ in this study compared to the previous studies. Firstly, with the exception of van Loon & van Dierendonck (2015, 2016), the scoring was done directly and no video footage was used. As mentioned earlier, scores derived from video footage can result in less detailed observations and seems inferior to real life observation. Secondly, the previous studies validating the EQUUS-FAP did not include ridden or trotting horses. Next to this, it is more difficult to completely remove bias from the live observations (if it is possible to distinguish the control horses and the horses experiencing pain. This could result in subconsciously giving higher scores to the horses experiencing pain. In this study, this effect will not be present as it is not possible to distinguish the videos of the horses at baseline and the induced horses.

Other studies have been performed on pain scores in ridden horses: using video footage, Dyson et al. (2018a) developed an ethogram of a pain scoring system in ridden horses and studied its application in the determination of the presence of musculoskeletal pain. In this study, several behaviours occurred significantly more in lame horses. Several of these behaviours are included in the EQUUS-FAP (position of the ears, eyelids and mouth opening). A retrospective study was done by (Dyson et al., 2018b) on video footage of lame horses, before and after resolving their lameness by analgesia. The sum of facial scores used in this study were significantly reduced. In Dyson et al 2018b, the significance of the separate parameters was determined. Significant facial parameters mentioned were: the mouth opening and shutting repeatedly, exposed sclera and changes in ear position. Both studies have several aspects in common with this study: they use video footage and score similar facial parameters. The second study also has a control group of nonlame horses. Therefore, it was expected that the induced Icelandic horses would score significantly higher than the baseline horses. The different outcome is possibly explained by the following differences: In the studies performed by Dyson et al (2018a, 2018b) it was not possible to hide the lameness of the horses. It is possible that the observers bias influenced the scoring. The video footage used by Dyson et al (2018a, 2018b) was significantly longer (3,2 +/- 0,3 minutes) and showed different gaits (trot and canter, 10 meter diameter circles) as opposed to the footage of the Icelandic horses being only 10 seconds long. The hoses were only filmed on a straight line in trot and tölt. The longer footage allows for a more extensive observation. This can result in noticing pain associated behaviours more accurately.

The total average score of the ridden horses in trot was significantly higher (p<0,001) than the total average scores of the trotting horses in hand for both induced and baseline horses. Using boxplots, we can see that the parameters that are most likely to score significantly higher for ridden horses are: the nostrils. the restless mouth and the mouth/lips/muscle tension.

At this moment, no studies have been published using a (facial parameter) pain scale, comparing ridden horses to horses in hand. Even though the absence of a significant difference in induced and baseline horses seems to contradict the EQUUS-FAP ICELANDs potential in recognising pain in these horses, this could indicate that the ridden horses experience more pain than horses in hand. This finding could possibly point towards a welfare issue affecting a big group of horses. It is important to know if the higher altered EQUUS-FAP pain scores actually indicate the ridden horses experience more pain as there are more possible explanations for this result.

It is possible that the opening of the mouth and the tension in the lips is always higher, because there is more tension on the bit when the horses are ridden. Especially as two of the parameters scoring higher in ridden horses include the restless mouth and the mouth/lips/muscle tension. In 2009 Manfredi et al., conducted a fluoroscopic study comparing the responses of horses to the presence of a bit and the effects of rein tension. Fluoroscopic images were made of the six horses used in this study, wearing 3 different kinds of bits, with loose reins and with rein tension. The horses receiving tension on the reins exhibited significantly more time retracting the tongue, mouthing the bit and bulging the tongue over the bit. These behaviours could all be resulting in higher scores on the parameters restless mouth and mouth/lips/muscle tension. It has to be noted that these horses were standing still and pressure on the reins was applied standing behind the horses. Eisersiö et al. (2013), investigated the behavioural responses of ridden horses at different head positions and different amounts of rein pressure. Using seven warmblood horses and riders on a high speed treadmill, video footage was collected. The data is not conclusive: while right rein tension increased the amount of mouth movement (as hypothesized), left rein tension decreased the amount of mouth movement. For an outcome to fully apply on the situation of this study, it would be ideal if both ridden horses and horses in hand were included. No prior study on the effect of riding on mouth movement is found meeting this exact criteria.

It would be possible to suggest that the higher weight (of a horse including a rider) results in more pressure on the shoes and therefore would cause more pain to the induced ridden horses. But it is important to note that the ridden horses in baseline have a higher score when ridden as well. The extra weight of the rider could result in the horse needing to put more effort into moving, leading to more opened nostrils and a higher score on this parameter. In a complementary study, it is advised to use a validated pain scale for ridden horses and horses in hand, including different gaits and riders with divergent levels of skills.

Conclusions

The usage of the EQUUS-FAP ICELAND in moving Icelandic horses (trot and tölt), experiencing orthopaedic pain due to induction of lameness, and in baseline shows promise but needs further research. The facial pain scale demonstrated a good inter-, and intra-observer reliability for both experienced veterinarians and unexperienced veterinary students. However, the scale was not able to differentiate between induced horses and horses at baseline (both the total average EQUUS-FAP ICELAND score as the scores of the separate parameters). For an additional study, it is advised to score the horses in real life instead of using video material, score for a longer period than 10 seconds and include different gaits. A significantly higher average total EQUUS-FAP ICELAND score was found in ridden horses than in horses that were trotted in hand, in horses with induced lameness and baseline horses. The parameters most likely to cause this significant difference are the nostrils, mouth/lips/muscle and restless mouth. With current data, it is not possible to certainly state that this difference is caused by the ridden horses experiencing more pain then the horses in hand. The rider seems to have an effect needs to be researched further.

Acknowledgments

I would like to thank my supervisors, J. P. A. M. van Loon and F. M. Serra Bragança for all of their efforts, kindness and their enormous amount of patience supervising this project. Thank you so much for providing me with all this feedback, the video material and scoring the videos with me. Secondly, I would also like to thank Katherina Ask and Iris Pluim for scoring a part of the footage. And lastly, I would also like to thank Hólar university in Iceland, and everyone involved in filming the used footage.

References

Ahola Kohut, S., Pillai Riddell, R., Flora, D. B., & Oster, H. (2012). A longitudinal analysis of the development of infant facial expressions in response to acute pain: Immediate and regulatory expressions. *Pain*, *153*(12), 2458–2465. https://doi.org/10.1016/j.pain.2012.09.005

Ashley, F. H., Waterman-Pearson, A. E., & Whay, H. R. (2005). Behavioural assessment of pain in horses and donkeys: Application to clinical practice and future studies. *Equine Veterinary Journal*, *37*(6), 565–575. https://doi.org/10.2746/042516405775314826

Bufalari, A., Adami, C., Angeli, G., & Short, C. E. (2007). Pain assessment in animals. *Veterinary Research Communications*, *31*(SUPPL. 1), 55–58. https://doi.org/10.1007/s11259-007-0084-6

Bussières, G., Jacques, C., Lainay, O., Beauchamp, G., Leblond, A., Cadoré, J. L., Desmaizières, L. M., Cuvelliez, S. G., & Troncy, E. (2008). Development of a composite orthopaedic pain scale in horses. *Research in Veterinary Science*, *85*(2), 294–306. https://doi.org/10.1016/J.RVSC.2007.10.011

Costa, E. D., Stucke, D., Dai, F., Minero, M., Leach, M. C., & Lebelt, D. (2016). Using the Horse Grimace Scale (HGS) to Assess Pain Associated with Acute Laminitis in Horses (Equus caballus). *Animals : An Open Access Journal from MDPI*, *6*(8). https://doi.org/10.3390/ANI6080047

Dalla Costa, E., Minero, M., Lebelt, D., Stucke, D., Canali, E., & Leach, M. C. (2014). Development of the Horse Grimace Scale (HGS) as a pain assessment tool in horses undergoing routine castration. *PLoS ONE*, *9*(3). https://doi.org/10.1371/journal.pone.0092281

de Grauw, J. C., & van Loon, J. P. A. M. (2016). Systematic pain assessment in horses. *Veterinary Journal*, *209*, 14–22. https://doi.org/10.1016/j.tvjl.2015.07.030

Dyson, S., Berger, J., Ellis, A. D., & Mullard, J. (2018a). Development of an ethogram for a pain scoring system in ridden horses and its application to determine the presence of musculoskeletal pain. *Journal of Veterinary Behavior*, *23*, 47–57. https://doi.org/10.1016/J.JVEB.2017.10.008

Dyson, S., Berger, J. M., Ellis, A. D., & Mullard, J. (2017). Can the presence of musculoskeletal pain be determined from the facial expressions of ridden horses (FEReq)? *Journal of Veterinary Behavior*, *19*, 78–89. https://doi.org/10.1016/J.JVEB.2017.03.005

Dyson, S., Berger, J. M., Ellis, A. D., & Mullard, J. (2018b). Behavioral observations and comparisons of nonlame horses and lame horses before and after resolution of lameness by diagnostic analgesia. *Journal of Veterinary Behavior*, *26*, 64–70. https://doi.org/10.1016/J.JVEB.2018.05.001

Eisersiö, M., Roepstorff, L., Weishaupt, M. A., & Egenvall, A. (2013). Movements of the horse's mouth in relation to horse–rider kinematic variables. *The Veterinary Journal*, *198*(SUPPL1), e33–e38. https://doi.org/10.1016/J.TVJL.2013.09.030

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/J.JCM.2016.02.012

Langford, D. J., Bailey, A. L., Chanda, M. L., Clarke, S. E., Drummond, T. E., Echols, S., Glick, S., Ingrao, J., Klassen-Ross, T., Lacroix-Fralish, M. L., Craig, K. D., & Mogil, J. S. (2010). Coding of facial expressions of pain in the laboratory mouse. *Nature Methods*, *7*(6), 447–449. https://doi.org/10.1038/nmeth.1455

Lindegaard, C., Thomsen, M. H., Larsen, S., & Andersen, P. H. (2010). Analgesic efficacy of intra-articular morphine in experimentally induced radiocarpal synovitis in horses. *Veterinary Anaesthesia and Analgesia*, *37*(2), 171–185. https://doi.org/10.1111/J.1467-2995.2009.00521.X

Manfredi, J. M., Rosenstein, D., Lanovaz, J. L., Nauwelaerts, S., & Clayton, H. M. (2009). Fluoroscopic study of oral behaviours in response to the presence of a bit and the effects of rein tension. *Comparative Exercise Physiology*, *6*(04), 143–148. https://doi.org/10.1017/S1755254010000036

Mullard, J., Berger, J. M., Ellis, A. D., & Dyson, S. (2017). Development of an ethogram to describe facial expressions in ridden horses (FEReq). *Journal of Veterinary Behavior*, *18*, 7–12. https://doi.org/10.1016/J.JVEB.2016.11.005

Price, J., Marques, J. M., Welsh, E. M., & Waran, N. K. (2002). Pilot epidemiological study of attitudes towards pain in horses. *Veterinary Record*, *151*(19), 570–575. https://doi.org/10.1136/vr.151.19.570

Sotocinal, S. G., Sorge, R. E., Zaloum, A., Tuttle, A. H., Martin, L. J., Wieskopf, J. S., Mapplebeck, J. C. S., Wei, P., Zhan, S., Zhang, S., King, O. D., & Mogil, J. S. (2011). The Rat Grimace Scale: A partially automated method for quantifying pain in the laboratory rat via facial expressions. *Molecular Pain*, *7*. https://doi.org/10.1186/1744-8069-7-55

Taylor, P. M., Pascoe, P. J., & Mama, K. R. (2002). Diagnosing and treating pain in the horse: Where are we today? *Veterinary Clinics of North America: Equine Practice*, *18*(1), 1–19. https://doi.org/10.1016/S0749-0739(02)00009-3

van Loon, J., Verhaar, N., van den Berg, E., Ross, S., & de Grauw, J. (2020). Objective assessment of acute pain in foals using a facial expression-based pain scale. *Animals*, *10*(9), 1–10. https://doi.org/10.3390/ani10091610

van Loon, Jonckheer-Sheehy, V. S. M., Back, W., René van Weeren, P., & Hellebrekers, L. J. (2014). Monitoring equine visceral pain with a composite pain scale score and correlation with survival after emergency gastrointestinal surgery. *The Veterinary Journal*, *200*(1), 109–115. https://doi.org/10.1016/J.TVJL.2014.01.003

van Loon, & van Dierendonck, M. C. (2015). Monitoring acute equine visceral pain with the Equine Utrecht University Scale for Composite Pain Assessment (EQUUS-COMPASS) and the Equine Utrecht University Scale for Facial Assessment of Pain (EQUUS-FAP): A scale-construction study. *The Veterinary Journal*, *206*(3), 356–364. https://doi.org/10.1016/J.TVJL.2015.08.023

van Loon, & van Dierendonck, M. C. (2017). Monitoring equine head-related pain with the Equine Utrecht University scale for facial assessment of pain (EQUUS-FAP). *Veterinary Journal*, *220*, 88–90. https://doi.org/10.1016/j.tvjl.2017.01.006

van Loon, & van Dierendonck, M. C. (2019). Pain assessment in horses after orthopaedic surgery and with orthopaedic trauma. *Veterinary Journal*, *246*, 85–91. https://doi.org/10.1016/j.tvjl.2019.02.001

VanDierendonck, M. C., & van Loon. (2016). Monitoring acute equine visceral pain with the Equine Utrecht University Scale for Composite Pain Assessment (EQUUS-COMPASS) and the Equine Utrecht University Scale for Facial Assessment of Pain (EQUUS-FAP): A validation study. *Veterinary Journal*, *216*, 175–177. https://doi.org/10.1016/j.tvjl.2016.08.004

Wagner, A. E. (2010). Effects of Stress on Pain in Horses and Incorporating Pain Scales for Equine Practice. *Veterinary Clinics of North America: Equine Practice*, *26*(3), 481–492. https://doi.org/10.1016/J.CVEQ.2010.07.001

Woolf, C. J. (1989). RECENT ADVANCES IN THE PATHOPHYSIOLOGY OF ACUTE PAIN. *British Journal of Anaesthesia*, *63*(2), 139–146. https://doi.org/10.1093/BJA/63.2.139