Validation of a Composite Pain Scale for Use with Horses Experiencing Acute Colic Pain

Marianne Beikes

Summary

Pain recognition and management in horses has become more important in recent years. Interest in the animal's welfare has increased. Horses are prey animals, and therefore will not easily show their pain. This study attempted to create and validate to objectively score and interpret pain in horses with acute colic. We have used an adapted Composite Pain Scale (CPS) system previously developed and tested by others on horses with orthopedic or post-operative somatic pain.

Eight horses with acute colic were scored using the colic CPS and these scores were compared to 18 healthy pain free horses (farrier clients and recipient mares). Two observers scored the animals, and scores were compared between observer 1 and observer 2 in order to determine the CPS's reliability. Videos of all horses were scored by other veterinarians using a Visual Analogue Scale (VAS) after which these scores were compared to the scores given using the CPS.

Animals not experiencing any pain were shown to score low, whereas horses with acute colic pain scored significantly higher on the CPS. Pain scores for colic patients decreased over time during their stay at the clinic. Scoring based on the CPS was significantly more reliable between observers than scoring performed with the VAS.

This study found that the CPS system could become a reliable method to use in the objective assessment of pain in horses with acute colic, and therefore benefit clinicians as well as owners and caretakers.

Introduction

The timely recognition and proper management of pain in animals has become increasingly important over the past years. Society and owners are showing more interest in the wellbeing of animals. Research helps us gain knowledge about the different types of pain in relation to diseases and trauma. It is impossible to ask the animal in question if it is in pain and how much pain it is experiencing, which makes the assessment of pain in animals very subjective in most cases (van Loon et al. 2010). In the wild, the horse is subject to predation. An individual horse showing obvious signs of pain would make an easy target for the predator, therefore a horse will not readily display any behavior that indicates discomfort.

Despite the animals not easily showing that something is wrong, evidence exists that horses experience pain in a way similar to humans; the pain becomes an emotional experience as well as a physical one (Taylor et al. 2002). This makes it important to assess pain that a particular horse is experiencing in an adequate and objective manner.

Human medicine has developed a pain VAS (Visual Analogue Scale). Based on a number between 0 and 10, patients give the doctors feedback about the amount of pain they are in (Hielm-Bjorkman et al. 2011). The system is now also used to assess pain in babies, young children, and the elderly. It was adopted to help score the pain in laboratory animals during animal experiments, and even tested to be used by dog owners to determine whether or not their dog was in pain. The dog's specific pain behavior was difficult to recognize for most owners, which made this system very unreliable for use by animal owners.

Pritchett (Pritchett et al. 2003) aimed to identify the behavioral elements a horse in pain might display in order to be able to develop an NRS (Numeric Rating Scale). It was developed specifically for horses that had undergone exploratory abdominal surgery. The study focused mainly on types of behavior. It became clear that horses in a post-operative situation spent a significantly larger amount of time on certain types of behavior such as standing still and having their ears faced back compared to horses that had not had any surgery and so were not expected to be in pain.

The next step in the development of a quantifiable pain scale came from Bussières (Bussieres et al. 2008). He expanded on Pritchett's system by adding physiological variables and more detailed descriptions of behavior, in which several of these behavioral elements were used to develop a Composite Orthopaedic Pain Scale. This scale consists of a table in which behavioral elements, physiological parameters and interactive parameters help the observer score the behavior a horse is displaying (Figure 1).

Response to treatment	Criteria	Score/06
Interactive behaviour	Pays attention to people	0
	Exaggerated response to auditory stimulus	1
	Excessive-to-aggressive response to auditory stimulus	2
	Stupor, prostration, no response to auditory stimulus	3
Response to palpation of the painful area	No reaction to palpation	0
	Mild reaction to palpation	1
	Resistance to palpation	2
	Violent reaction to palpation	3
Behaviour	**Criteria**	**Score/21**
Appearance (reluctance to move, restlessness, agitation and anxiety)	Bright, lowered head and ears, no reluctance to move	0
	Bright and alert, occasional head movements, no reluctance to move	1
	Restlessness, pricked up ears, abnormal facial expressions, dilated pupils	2
	Excited, continuous body movements, abnormal facial expression	3
Sweating	No obvious signs of sweat	0
	Damp to the touch	1
	Wet to the touch, beads of sweat are apparent over the horse's body	2
	Excessive sweating, beads of water running off the animal	3

Figure 1. Example of two behavioral elements as used in Bussières' study (Bussieres et al. 2008).

Each element is linked to a number value and the summed score in points represents the "pain score" of that horse at that moment. The higher the total amount of points scored, the more pain a horse is experiencing. Even though not all of the elements Bussières et al. describe in their CPS can be translated to visceral pain, a few of them can still be used for patients with acute abdominal pain(van Loon et al. 2010).

These elements formed the basis for the pilot study by van Loon et al.(2010) that precedes the current study. This pilot study was performed observing a wide variety of clinical patients admitted for a large number of different reasons.

Despite the fact that Pritchett (Pritchett et al. 2003) have studied behavior that could indicate a horse is experiencing visceral pain, all cases in this study were scored in a post-operative situation following emergency abdominal surgery and therefore the results can not be extrapolated to acute colic(Pritchett et al. 2003). Sellon et al.(Sellon et al. 2004) and Sanz et al.(Sanz et al. 2009) have used Pritchett's system. In these studies the main focus was also post-operative pain. Despite the pilot study in which van Loon et al.(2010) assessed several types of pain, acute colic pain was not included(van Loon et al. 2010).

Pritchett et al.(Pritchett et al. 2003) and later Sellon et al.(Sellon et al. 2004) and Sanz et al.(Sanz et al. 2009) all used the very concise composite NRS developed by Pritchett(Pritchett et al. 2003). This NRS contained only a few behavioral elements and each element was given a score between 1 and 4. The behavior descriptions in the NRS were brief. Bussières et al. were the first to use a more detailed multidimensional system(Bussieres et al. 2008) describing many more elements and also expanding on the existing descriptions. This was likely the reason van Loon et al.(van Loon et al. 2010) and Graubner et al.(Graubner et al. 2011) decided to adopt this system instead of Pritchett's. It makes sense that a more complex system is preferable, especially with accurate descriptions of the behavior that may be observed.

The use of more detailed, less ambiguous behavioral and physiological elements in a pain scale will leave less to the observer's interpretation and will therefore make the scoring of the patient's behavior more objective and much easier to reproduce.

The average clinician will meet many patients with acute abdominal pain and so will be faced with many situations where pain management in the acute stages becomes an important concern. It is difficult to judge at first glance how much pain an individual horse is experiencing, especially because the initial focus at this point is on diagnosing the cause, not on diagnosing the pain.

Currently, there is no validated system available that every veterinarian, caretaker and perhaps even owner can use to objectively assess pain in a horse during the acute stages of colic. In many cases it is even very difficult to objectively assess whether the pain is increasing or decreasing over time.

The goal of this study is to develop and validate a Composite Pain Scale specifically tailored to fit patients with symptoms of acute abdominal pain. Using the CPS that van Loon et al. (2010) described, which in itself is an adaptation of Bussières orthopaedic system(Bussieres et al. 2008), as a base, this study will attempt to develop a system that can be used by clinicians as well as caretakers and perhaps owners to objectively assess acute visceral pain in a patient.

Hopefully, with this system, decisions about pain management will be easier to make and the outcome more accurate. For this study horses that arrived at the veterinary hospital experiencing any form of acute abdominal pain were assessed with a new pain scale adapted from the system used by Bussières et al. (2008) and van Loon et al. (van Loon et al. 2010). Several control groups were determined and assessed through use of the same system. The outcomes were then compared with usage of the VAS system on videos of the same horses, performed by different groups of clinicians.

This study was performed in order to research the possibility of creating a reliable method of scoring pain in horses with acute colic. The aim is to have an easy to use, easy to reproduce, valid system that aids in the assessment of pain in these cases. In addition, the possibility of creating a similar system to be used on the horse's facial expression was researched.

The expected outcome is a system that will be found reliable through observation of several different and accurately described behavioral, interactive and physiological parameters. It can then be shaped into a system that is also easy and quick to use.

Materials and methods

Animals

Adult horses admitted to the Department of Equine Sciences of the Faculty of Veterinary Medicine in Utrecht, the Netherlands, between March and June 2013 were used for this study (Table 1). The patient group consisted of eight horses of both genders presented to the clinicians with abdominal pain (mean age 11 years and 9 months; 5 Dutch Warmbloods, 1 Gypsy Vanner, 2 mixed breeds). The control group consisted of eighteen horses that were free of pain before and during the test. Of this control group, ten horses were healthy faculty owned recipient mares for embryo transfer (mean age 5 years and 2 months; 8 Dutch Warmbloods, 1 Belgian Warmblood, 1 mixed breed) . The other eight were horses of both genders coming in for an appointment with the farriers and judged not to be in any type of pain, including orthopaedic pain (mean age 10 years; 5 Dutch Warmbloods, 2 mixed breeds and 1 Arabian thoroughbred).

Owners were informed about the procedure beforehand and asked to sign a form of consent as well as fill out a short questionnaire about their animal's daily routine (feeding, exercise) and housing (see Appendix A). They were allowed to watch during the procedure if they so wished as long as they remained relatively silent.

Table 1. An overview of the horses used during the study, including the number of observations made.

Group	Number of horses	Total number of observations	Paired observations*	Single observations**
Colic patients	8	28	11	17
Farrier clients	8	16	16	0
Embryo transplant mares	10	20	20	0
Total	26	64	47	17

*Observations made by two observers at once.
**Observations made by one observer.

Composite Pain Scale

For this study, we used the Composite Pain Scale originally developed by Bussières et al.(Bussieres et al. 2008) as a starting point, which was tested by van Loon et al.(van Loon et al. 2010) to study its practical application and suitability to asses visceral pain. In the underlying study this scale was further adapted to specifically help us assess acute abdominal pain. Specific behavior relating to orthopaedic pain such as weight distribution on the feet was therefore removed from the CPS in our adaptation. Furthermore, since colic patients are not allowed to eat hay until later on, we also removed this element. Elements such as looking at the flank, (excessive) pawing on the floor and kicking at the abdomen were copied directly since this type of behavior is also displayed by horses with acute abdominal pain. With the help of an expert in equine behavior (Machteld van Dierendonck) we added elements that are believed to reflect abdominal pain such as lying down, rolling and auditory expressions of pain (moaning, teeth grinding). Because pain may lead to irritation, excessive tail flicking was also added to our scale.

Our pain scale is a composite numeric rating scale (NRS) containing physiological parameters, behavioral elements and interactive elements(Appendix B).

Behavior was split into interactive elements and behavior displayed when no humans are present in the stable. Each element contained an NRS score between 0 and 3. The single NRS scores for each element were summarized to a total score ranging from 0 (no pain) to 42 (severest pain imaginable).

Facial Expression Pain Scale (FEPS)

One of the observers also developed a system similar to the CPS system in order to determine whether or not the horse's facial expression adequately conveys pain (Appendix C). Observing the horse's face was done separately, directly following the observation of the horse in the stall.

Visual Analogue Scale (VAS)

After the observational period a group of volunteers was asked to assess the pain in each horse using the Visual Analogue Scale (VAS) as they would do if this was their patient. This group consisted of two veterinarians, two caretakers, two veterinary master students and two veterinary bachelor students. Each observer was given a one minute video of an observation and asked to score the pain they believed the individual horse was in with a number ranging from 0 (no pain) to 10 (severest pain imaginable). All of the volunteers were blinded to the cases. A total of six videos were presented to the volunteers, carefully selected to display the range of patients and control horses included in the study.

Two interns, also blinded to the cases, were asked to independently score all of the observations made throughout the observational period. As with the other volunteers, they were requested to use the VAS system and assess the horse in each video as if it was their own patient.

Temperament testing

A temperament test was developed with the help of equine behavior expert Machteld van Dierendonck and a study on the subject, conducted by Hausberger et al.(Hausberger et al., 2009). The test itself was not executed due to time restrictions, but can be viewed in Appendix D.

Experimental setup

When possible, pain scoring was performed by two students separately at the same time. However, these students were aware of each individual horse's case when admitted to the experiment and thus not blinded for the duration of the study. These scores were used to calculate the interobserver variability. Before the start of the observational period, the CPS system was tested on horses not included in the study.

Care was taken not to intervene with the veterinarians' normal routine. If possible, each colic horse was observed immediately after arrival or at the earliest convenient time after the initial checkup. Each horse was placed in a stall specially designed for colic patients and given a mask that prevented the horse from eating the bedding, regardless of their health status. The horse was then observed and scored for 5 consecutive minutes.

The control group was observed in the same way. Farrier clients were all observed after receiving their new shoes. Embryo transplant recipient mares were taken from their own stable and put directly into the colic stall. All horses were allowed five minutes to acclimate themselves to their new environment if necessary before the start of the observation.

After the initial observation and scoring, the observers entered the stable in order to gather the physiological data and then made a minute long observation of the head for the FEPS.

Control group horses were observed only once at random times of day (depending on their appointments). Patient horses were observed several times. The aim was to assess them at the time of admittance (variable), the first morning after admittance and the second morning after admittance. If it was decided to administer pain medication at the clinic after arrival, the second observation took place one hour after medication was given. In cases where the horse had received

analgesia before being transported to the clinic or the horse's pain management status was unknown, the horse was observed again two to three hours after arrival. Horses requiring surgery after the initial diagnosis were not observed afterwards.
Videos were made of each horse while it was in the colic stable with a camera that was attached near the ceiling in the front left corner of the stable (Figure 2). Attached to the camera were a video screen and a recording device to save the videos for later use.

Figure 2. Example of a horse in the colic stable and the angle at which each horse was filmed. In the bottom right corner the observers' documents can be seen: this was the angle from which the two observers assessed each horse.

Data and statistical analysis

Analysis of the gathered data was done using Microsoft Excel 2007 and SPSS. The data were not tested for normality. Due to the small number of horses, it was assumed that there could be no normal distribution until more data are obtained. For interobserver reliability and comparing the CPS results to VAS results, the correlation coefficient was calculated using a Spearman's correlation analysis. An ANOVA comparison was used to compare each group (patients, farrier clients and embryo transplant mares) in order to determine whether there was a significant difference in scores. The Mann-Whitney U test was used to compare the patient group to the control group and to compare conservative treatment patients to patients that received surgery or were euthanized. Significance levels of $p < 0.05$* were used for all statistical analyses.
In this study, the data gathered from the questionnaires for the owners were not statistically analyzed. All of the answers were documented and saved for use in follow-up studies.

Results

Observations made

A first observation was successfully completed for 28 horses. Of the 8 patient horses, 4 could be observed until the second morning of their stay. One horse was observed three hours after arrival at the clinic because it was unclear whether or not the horse had received pain medication before transport. None of the patient horses were assessed one hour after arrival. In most cases, the first observation of a colic horse was carried out by only one observer because many of the patients arrived very late at night. It was agreed that during nights only one observer would be present in order to avoid fatigue.

ANOVA between groups

A one-way ANOVA was performed in order to determine whether the farrier horses and embryo transfer mares could be considered one control group, compared to the colic patient group. Table 2 shows that there is no significant difference between the farrier and embryo transfer mares, but there is a significance between either of these groups and the patient group. Thus in the following analyses, both farrier horses and recipient mares were used as one group.

Table 2. Outcome of one-way ANOVA: multiple comparisons using Bonferroni. ET mares = Embryo transfer recipient mares. F = 6.606; df = 2.

(I) sub group	(J) sub group	Sig.
Farriers	ET mares	1,000
	Colic	,010**
ET mares	Farriers	1,000
	Colic	,017*

* Significance $p < 0.05$

** Significance $p < 0.01$

Sensitivity of the CPS

The box-and-whiskers graph shows the scores given by the two main observers to colic patients and healthy horses (Figure 3). The Mann-Whitney U test resulted in a $p < 0.001$.

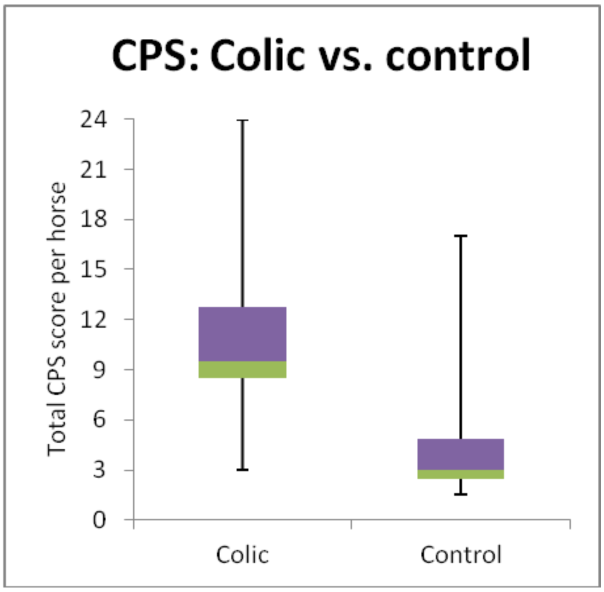

Figure 3. Box-and-whiskers comparing colic patients (n = 8) to healthy horses (n = 18). The box represents 50% of the gathered data. The whiskers indicate the lowest and highest scores given by observers. The border between green and purple shows the median score***.

In order to test sensitivity, two categories of patients were compared: a conservative treatment group and a group in which surgery or euthanasia was recommended. These two groups were then compared using a Mann-Whitney U test (Figure 4). These groups showed a tendency toward significantly different CPS scores ($p = 0.071$).

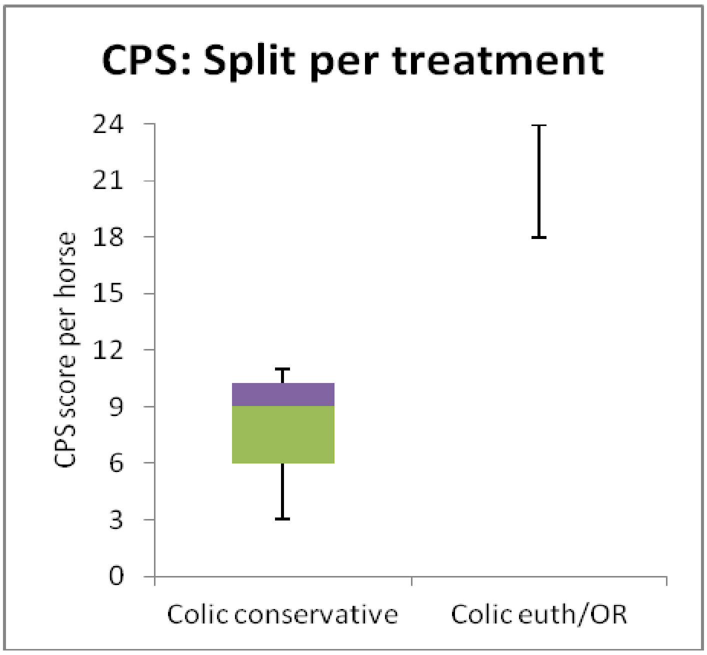

Figure 4. Comparison of horses that received conservative treatment (colic conservative) after admittance (n = 6) and horses for which surgery or euthanasia was advised (colic euth/OR) (n = 2).

To determine whether the system shows the effect of time on the individual horse's pain, the patients which were followed over time were also compared (Figure 5). No statistical analysis was performed on these data due to a low n.

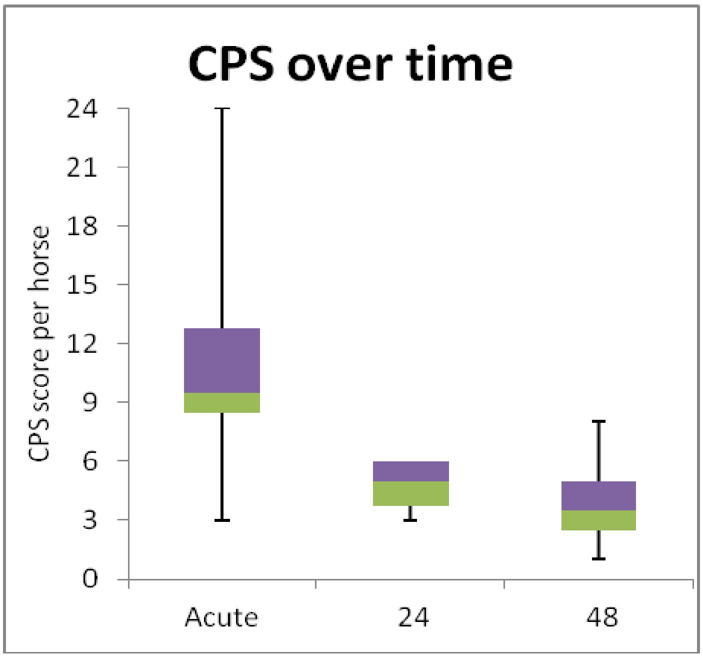

Figure 5. Box-and-whiskers graph giving an impression of the scores in time (hours after admittance): acute colic at admittance (n = 8), 24 hours after admittance (n = 4) and 48 hours after admittance (n = 4).

Interobserver reliability CPS and VAS

In order to test the system's reproducibility the scores of the two main observers were compared using a Spearman's correlation analysis (Figure 6). The interobserver scores of the two independent observers showed a correlation coefficient of Rs = 0.942; n = 29 scores.

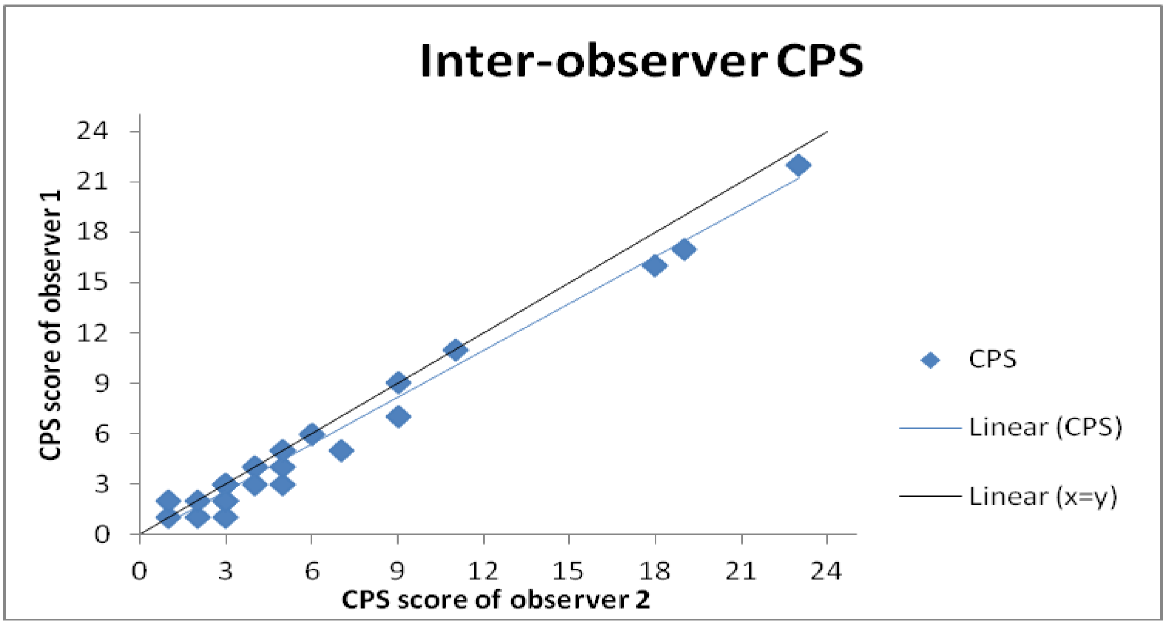

Figure 6. Interobserver reliability with n = 29 scores (all at t = 0 hours). The black line represents the ideal line x = y. The blue line is best fit to the gathered data. p < 0.001***.

VAS scores between the two independent observers A and B (both interns) resulted in a correlation coefficient of Rs = 0.166 with n = 29 (p = 0.389) (Figure 7).

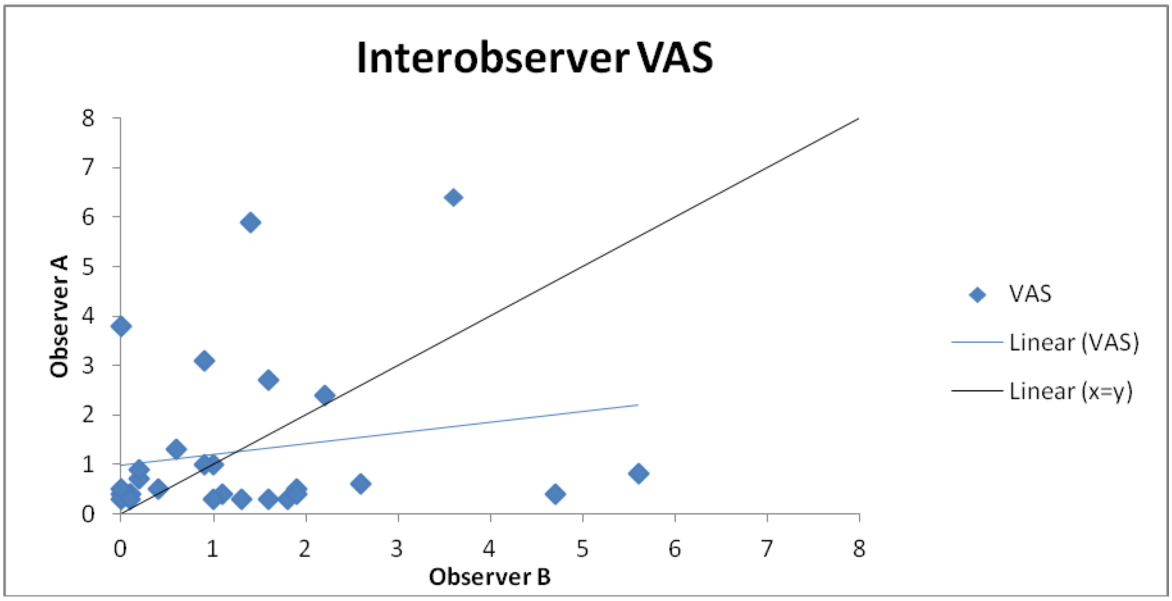

Figure 7. Interobserver reliability between observers A and B (interns) with n = 29 scores (all at t = 0 hours). The black line represents the ideal line x = y. The blue line is best fit to the gathered data.

Interobserver reliability of the VAS

In order to determine the interobserver reliability within each group of observers using the VAS, an ANOVA analysis was performed on the scores given to the six selected videos (table 3). These results show that the interns have a low interobserver reliability, but the other groups have a higher interobserver reliability.

Table 3. Outcomes of a two-tailed ANOVA showing the correlation and p-values within each group. * Indicates significant correlation: the two observers have given significantly similar scores, meaning only the intern pair has a very low interobserver reliability.

Group (n = 2 per group)	Correlation coefficient (Rs)	p value
Interns	0.406	0.425
Veterinarians	0.886	0.042*
Caretakers	0.829	0.042*
Master students	0.886	0.019**
Bachelor students	0.829	0.042*

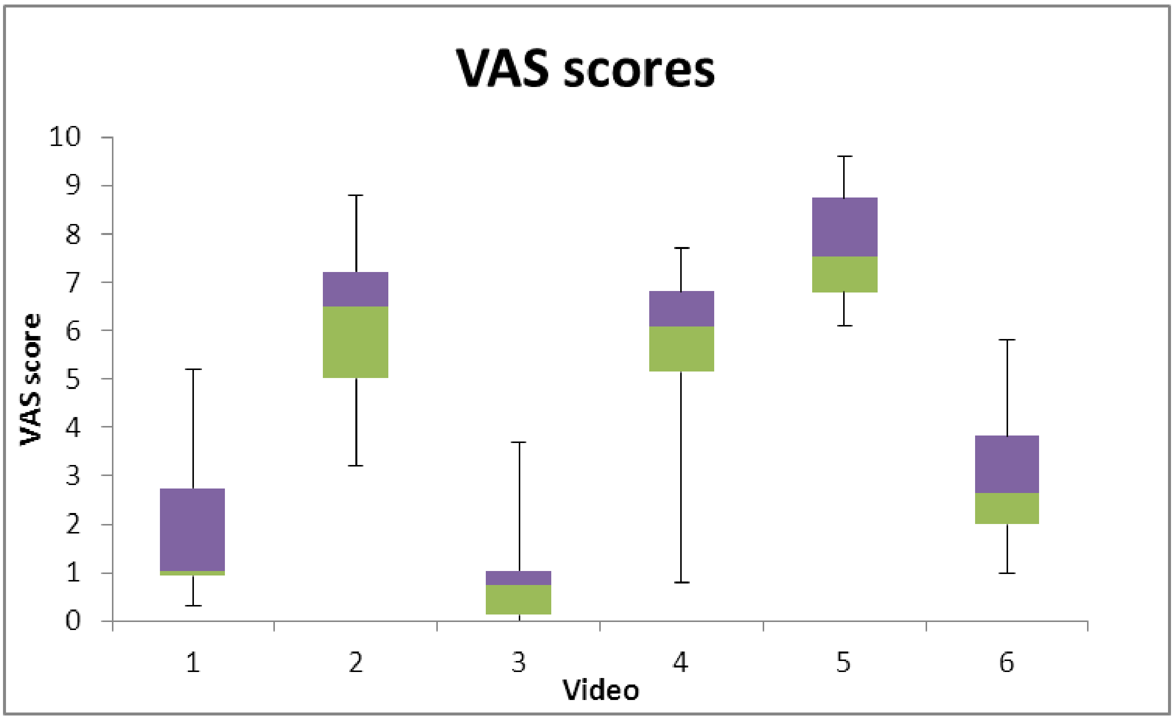

Figure 8. Variations in the scores given by pairs of observers using the VAS system for the six selected videos. The average score of each pair was used. The “interns” group also scored all the other videos as seen in Figure 7.

CPS versus VAS

The CPS (by observers 1 and 2) and the VAS (by observers A and B) were compared using Spearman’s correlation analysis, resulting in a correlation coefficient of Rs = 0.331, n = 29, with a significance of p = 0.052 (Figure 9).

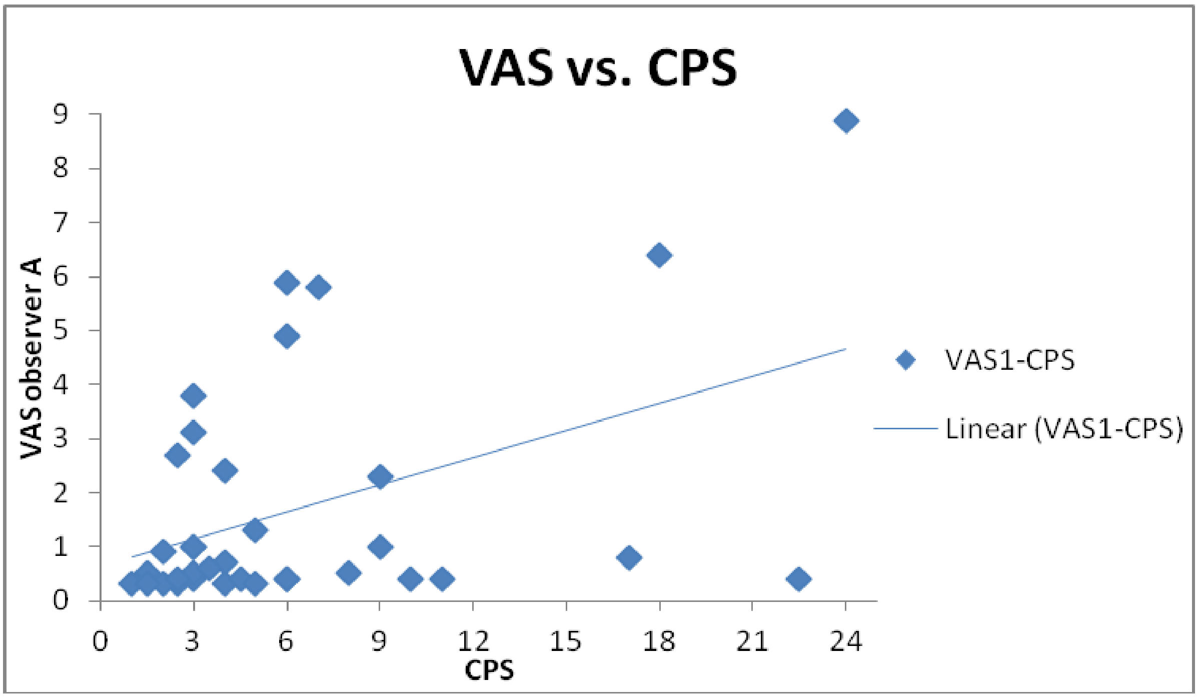

Figure 9. Comparison of CPS scores and VAS scores. The VAS of observer A was used as if it were an average of both VAS observers (A and B) only to plot this graph. The average CPS score of both observers 1 and 2 was used.

Lactate and pain scores

Lactate values paired with pain scores given using the CPS (Figure 10) were compared. The graph includes all eight colic horses.

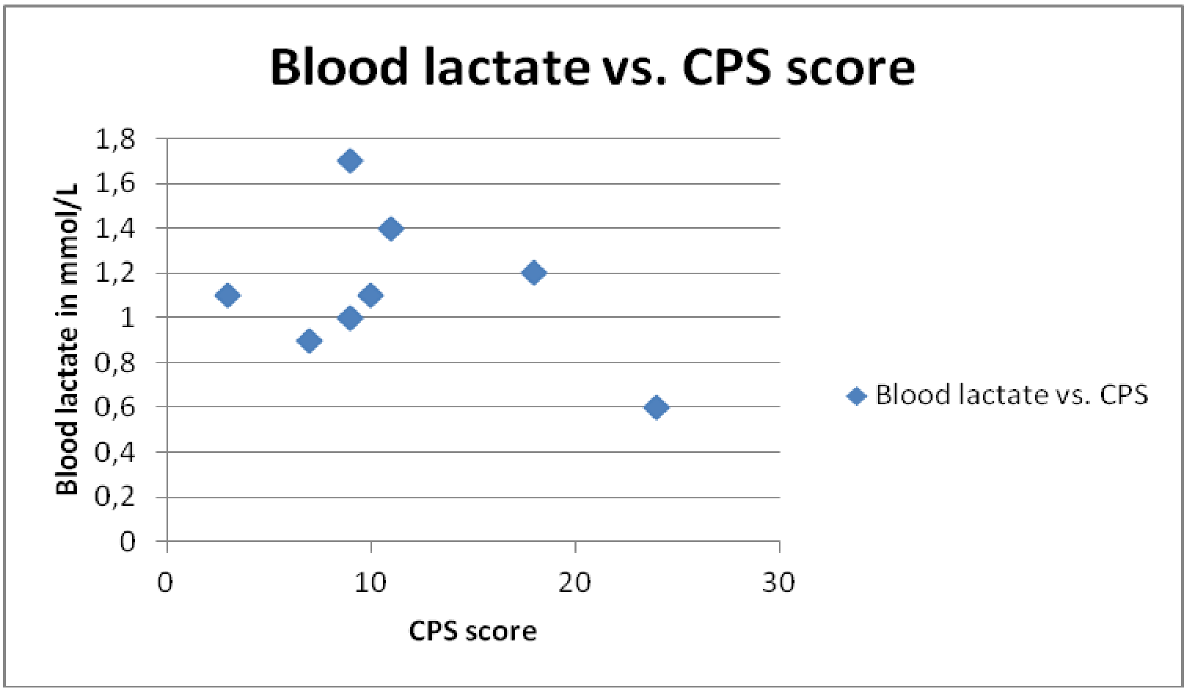

Figure 10. Scatter plot comparing blood lactate values at admittance and pain scores in colic horses in the acute phase.

Discussion

This study is the first to focus on pain in horses with acute colic. Though many types of pain have been studied(Pritchett et al. 2003, Sellon et al. 2004, Sanz et al. 2009, Bussieres et al. 2008, Graubner et al. 2011, van Loon et al. 2010b) in a similar manner, the pain experienced during acute colic was never observed very specifically. All previous studies focused on post-operative pain in horses.

The results show that the CPS used in this study has a high interobserver reliability, with consistent scores between both observers scoring independently. Because of this, it is more likely that different observers will always score the same amount for an individual horse, removing the possibility of observer bias influencing the score. It should be noted here that both observers were not always present at the same time to observe the most painful colic patients. However, for the horses where both observers could attend the scores were almost identical. In future studies, scoring will have to be done in pairs more often to ascertain the system's reliability even for the higher pain scores.

The scoring appears to be reliable irrespective of observer experience, even though Bussières et al.(Bussieres et al. 2008) and van Loon et al.(van Loon et al. 2010b) used comparable systems and also showed a similarly high interobserver reliability. Their studies used a system that focused on objective pain assessment in horses with experimentally induced acute inflammatory synovitis (Bussieres et al. 2008), and tailored their CPS to reflect elements that would indicate acute orthopaedic pain. Therefore for this study, their interobserver reliability did not tell us whether or not a CPS designed for acute colic pain could be reliable. Despite the small size of this study's population, the results already give a strong suggestion that we are on the right path to develop a reliable, easy to use scoring system.

The people who were using the Visual Analogue Scale (VAS) to score pain in all or several horses had good interobserver reliability between the pairs. However, it appears there was a difference between each group, especially between professionals (more experience) and students (less experience) (Table 3 and figure 8). The experienced people had a lower interobserver reliability and on average, tended to give lower scores. The less experienced (student) observers showed a tendency to assign higher scores to the horses. This suggests that experience could be of great influence when using the VAS to score pain in horses, which makes VAS a less reliable means for assessment of pain in horses with acute colic.

It is our belief that having a more detailed system incorporating many behavioral and physiological elements will create a more reliable system, removing the observer bias from the score. This is demonstrated by the differences in interobserver reliability between our CPS and the ultimately subjective VAS. Pritchett et al. (Pritchett et al. 2003) did a very brief analysis of interobserver reliability with an outcome measure in minutes' difference between observer's results. Graubner et al. (Graubner et al. 2011) applied the same methods as performed in this study but the outcome was low. The CPS developed by Bussières et al. (Bussieres et al. 2008) and also used by Van Loon et al. (van Loon et al. 2010b) both resulted in a correlation coefficient of about 0.8.

The adaptations we have made to the existing CPS first used by Bussières et al(Bussieres et al. 2008), then directly copied and used by van Loon et al.(van Loon et al. 2010) are important to note. By adding behavioral elements associated with colic pain and removing the elements associated with orthopaedic pain, the outcomes became a lot more reliable as demonstrated by the high correlation coefficient between observers 1 and 2.

Currently, most clinicians in human medicine as well as veterinary medicine assess pain in patients who cannot speak by means comparable to the VAS(Hielm-Bjorkman et al. 2011): a quick critical observation of the patient tells them whether or not they are in pain. This VAS system was used as a "golden standard" in this study representing the current method of assessing equine pain during acute colic, to determine whether or not the new method (the CPS) is better than the one currently in use (the VAS).

Many studies took blood samples in order to keep track of the plasma cortisol levels(Sellon et al. 2004, Bussieres et al. 2008, Sanz et al. 2009, Pritchett et al. 2003), but we chose not to do this. Cortisol indicates a higher stress level in the animal, but none of the studies showed whether or not plasma cortisol was a viable parameter to indicate pain in a horse. All of the studies mentioned that plasma cortisol levels were influenced by many factors other than just pain. In addition, we have chosen not to interfere too much with the patients' treatment or break the skin of healthy horses and therefore did not draw any blood. Another reason not to take this factor into account is that we aim to develop a scale that can be used in real time situations. Having to await laboratory testing would slow the observation down immensely.

We also chose not to use a time budget type of observation as used in Pritchett et al.'s (2003) NRS system(Pritchett et al. 2003) (example: "time spent pacing the stable"), because this is not relevant for an immediate assessment and would lengthen the process, rendering it useless for our goals.

Since the subjects in this study were dependent on incoming patients, there was no control over the ages and breeds used in this study. The studied horses had a mean age of 8.5 years, with 18 Dutch Warmblood horses, 5 mixed breeds, 1 Arabian thoroughbred, 1 Belgian Warmblood horse, and 1 Gypsy Vanner. The expectation is that the group will be more diverse when more patients take part in the study in the future.

There are several reasons for not including foals in this study. Young horses with colic symptoms require a lot of attention at this early stage in their lives. We deemed it unnecessarily inconvenient to interfere with this intensive treatment. We also considered it stressful to both mare and owner to observe young foals during their illness. A foal has a base heart rate, respiratory frequency and temperature that are different from an adult horse, so these parameters would have to be adjusted for the CPS to work on this age group. Additionally, the behavior of a preweanling horse is heavily influenced by its age and mother. The foal may react differently to outside stimuli (auditory and visual), may attempt to hide behind the mother or be scared of unfamiliar human contact(Lansade, Bouissou & Boivin 2007).

One of the healthy mares used in the control group had an exceptionally high score despite not being in pain. This mare was considered a "false positive": her video was also included in the 6 videos scored by secondary observers. Because of the small number of videos scored in this part of the study, a false positive has more weight in the statistical analyses. The influence of a false positive individual decreases with an increasing number of animals studied: this mare's score did not influence the outcome when included in the entire study population.

Due to the time available for this study, VAS observers were not asked to repeat the exercise at a later date. In order to be able to determine the intra-observer reliability using the VAS system, it can be beneficial to do so in the future. It is unlikely that the difference between the outcomes of the interns could be based on experience in using the CPS system. This is due to both interns that scored all of the videos having a nearly similar amount of experience with equine behavior. An alternative hypothesis to explain this difference could be the observer's personality. In this study, observers were not asked for their motivation. Perhaps it is useful to do so in future studies on this subject in order to determine what is causing these subjective differences.

It was impossible to do any statistic analyses on the different categories of treatment of the colic horses, due to not having a large enough patient group to make the outcome reliable. For future research it could be beneficial to compare scores for each group (conservative treatment versus euthanasia versus surgery).

One of the physiological elements considered for future addition to the CPS system was blood lactate. Therefore, throughout this study we gathered each patient's blood lactate value, which was measured by the attending veterinarian as a standard part of each colic patient's checkup. However, due to the small number of patients we were able to include in this study, it was not possible to do any statistical analysis on the blood lactate values versus the pain scores. When more data are added to the current study, perhaps a conclusion can be drawn about the value of blood lactate as a predictive parameter for the amount of pain an individual horse is experiencing, or even the prognosis. In this study, it was not possible to determine whether lactate in the blood accurately predicted the cause, the severity of the colic symptoms or the resulting pain, due to having observed a small number of patients.

Despite the indications that our method is useful for pain assessment in equine patients experiencing acute colic pain, it remains relatively uncertain how this study's CPS measures pain in relation to excitement or fear. In order to decrease the chance of misjudgement, it is useful to conduct further study in which scores are given before and after administration of pain medication and then compared to one another.

Differences in temperament can also be considered as a factor in the display of behavior(Hausberger et al. 2009). Practicality did not allow for each horse's temperament to be tested, but a concept for a test was already developed (Appendix D). For instance, temperament could be tested immediately after the CPS assessment of the control horses and on the second morning for patient horses. A second temperament test to be performed at home some time after the horse's admission to the hospital can also be considered. The results could be used to develop corrections for the eventual pain scale depending on the horse's temperament.

The results of the Facial Expression Pain Scale study are not included in this study, but can be found in the second observer's study report (Vermin, L. 2013).

In this first phase of the research, the control group was much larger than the group of colic patients we have gathered so far. Any comparisons made are therefore uncertain, though predictions for future results and suggestions for adaptations can be made.

Conclusion

The Composite Pain Scale has shown to have a very high interobserver reliability (r^2 of 0.94) compared to the lower reliability of using a Visual Analogue Scale (r^2 of 0.17). These results, despite being based on a lower number of patients, suggest that using a more complex multidimensional scoring system such as the CPS is more reliable and less subjective than the VAS method.

Ranges and average scores also differed between CPS users and VAS users. The CPS scores showed a smaller range and the averages were very close to the individual observers' scores. The VAS showed a larger range in scores with more deviations from the regression line compared to the CPS scores. Therefore it can be suggested that based on these results the CPS method seems much more reliable for use in consecutive cases and much easier to reproduce.

As can be expected, the CPS method distinguished between mild cases and severe colic. However, this conclusion is somewhat unreliable as of yet, due to the small amount of patients observed in this study.

No trend could be discovered in blood lactate compared to pain scores. In most cases a higher pain score also correlated with a higher lactate concentration, but this was not the case for all of the 8 horses observed.
Our recommendations are to continue gathering data for a follow-up study, so that a larger group of patients and control horses may make the statistical analysis of this system more accurate. In addition, if secondary observers are used to compare the CPS to the VAS again, it may be beneficial to ask them to motivate their scores.
In future studies we also recommend emphasizing attempts to score as many of the colic patients in the acute phase with two observers, since that is the most important score needed to validate this system.
For now, this study strongly indicates that the CPS can be a reliable method of scoring pain in patients objectively. If further optimized and developed into an easy to use standard, this system could greatly benefit clinicians and caretakers in the process of pain management.

References

Bussieres, G., Jacques, C., Lainay, O., Beauchamp, G., Leblond, A., Cadore, J.L., Desmaizieres, L.M., Cuvelliez, S.G. & Troncy, E. 2008, "Development of a composite orthopaedic pain scale in horses", *Research in veterinary science,* vol. 85, no. 2, pp. 294-306.

Graubner, C., Gerber, V., Doherr, M. & Spadavecchia, C. 2011, "Clinical application and reliability of a post abdominal surgery pain assessment scale (PASPAS) in horses", *Veterinary journal (London, England : 1997),* vol. 188, no. 2, pp. 178-183.

Hausberger,Martine, Roche,Hélène, Henry,Séverine & Visser, E.,Kathalijne, 2009 .

Hielm-Bjorkman, A.K., Kapatkin, A.S. & Rita, H.J. 2011, "Reliability and validity of a visual analogue scale used by owners to measure chronic pain attributable to osteoarthritis in their dogs", *American Journal of Veterinary Research,* vol. 72, no. 5, pp. 601-607.

Lansade, L., Bouissou, M.F. & Boivin, X. 2007, "Temperament in preweanling horses: development of reactions to humans and novelty, and startle responses", *Developmental psychobiology,* vol. 49, no. 5, pp. 501-513.

Pritchett, L.C., Ulibarri, C., Roberts, M.C., Schneider, R.K. & Sellon, D.C. 2003, "Identification of potential physiological and behavioral indicators of postoperative pain in horses after exploratory celiotomy for colic", *Applied Animal Behaviour Science,* vol. 80, no. 1, pp. 31-43.

Sanz, M.G., Sellon, D.C., Cary, J.A., Hines, M.T. & Farnsworth, K.D. 2009, "Analgesic effects of butorphanol tartrate and phenylbutazone administered alone and in combination in young horses undergoing routine castration", *Journal of the American Veterinary Medical Association,* vol. 235, no. 10, pp. 1194-1203.

Sellon, D.C., Roberts, M.C., Blikslager, A.T., Ulibarri, C. & Papich, M.G. 2004, "Effects of continuous rate intravenous infusion of butorphanol on physiologic and outcome variables in horses after celiotomy", *Journal of veterinary internal medicine / American College of Veterinary Internal Medicine,* vol. 18, no. 4, pp. 555-563.

Taylor, P.M., Pascoe, P.J. & Mama, K.R. 2002, "Diagnosing and treating pain in the horse. Where are we today?", *The Veterinary clinics of North America.Equine practice,* vol. 18, no. 1, pp. 1-19, v.

van Loon, J.P.A.M., Back, W., Hellebrekers, L.J. & van Weeren, P.R. 2010, "Application of a Composite Pain Scale to Objectively Monitor Horses with Somatic and Visceral Pain under Hospital Conditions", *Journal of Equine Veterinary Science,* vol. 30, no. 11, pp. 641-649.

Vermin, Larissa 2013, *Study report: Development and Validation of a Facial Expression Pain Scale*.

Appendix A: Questionnaire for patient owners

Name:
Horse's name:

General
Does your horse have any stable vices, ie. wood chewing, cribbing?

Your horse's temperament: introverted, not sensitive / average / extraverted, sensitive

Feeding schedule
Concentrates: ... x per day and ... kg per feeding
Roughage: ... x per day and ... kg per feeding
Have there been any changes in feed or feeding schedule recently? If yes, which?

Housing
Is your horse ever alone: never / occasionally / often

Stable
Type of stable: stall / free range stable / other:
Does your horse have any contact with other horses:
can see / can hear / can touch / shares stable with other horses / no contact at all
Bedding: flax / woodchips / straw / other:
Hours in stall per day:

Pasture: yes/no
Hours per day:
Number of animals in pasture:

Paddock/bak
Hours per day:
Is the horse given extra nutrition? If yes, what is given?

Is it possible that your horse ingests sand (during feeding or by purposely eating sand)?

Exercise
Dressage/show jumping/recreation/other:
Level:
Number of training sessions per week:
Average training time per session:

Appendix B: Composite Pain Scale used in this study

Date:
Time:
Observer:

Observation:
0 hours
~ 1 hour   if given NSAIDs at clinic
2 – 3 hours   if pain management is unknown or was administered before arrival at clinic → optional
first morning
second morning

Data	Categories	Score
Physiological data Heart rate	24 - 44 beats/min 45 - 52 beats/min 53 - 60 beats/min > 60 beats/min	0 1 2 3
Respiratory rate	8 - 13 breaths/min 14 - 16 breaths/min 17 - 18 breaths/min > 18 breaths/min	0 1 2 3
Rectal temperature	36.9 °C - 38.5 °C 36.4 °C - 36.9 °C or 38.5 °C - 39.0 °C 35.9 °C - 36.4 °C or 39.0 °C - 39.5 °C 35.4 °C - 35.9 °C or 39.5 °C - 40.0 °C	0 1 2 3
Digestive sounds	Normal motility Decreased motility No motility Hypermotility or steelband	0 1 2 3
Behavior Posture	Quietly standing and/or one hind leg resting, explores environment Slightly tucked up abdomen, still explores environment (with possible unrest) Extremely tucked up abdomen, hunched back and/or stretching of body/limbs Does not stand or for short amounts of time (<1 min), sits on hindquarters	0 1 2 3
Laying down, rolling	Does not lie down or rests lying down Lies down in normal posture, rolls or tries to roll (1-2 times/5 min) Alternates lying down and standing, rolls or tries to roll (>2 times/5 min) Constantly lies in an abnormal position: on its side with stretched limbs, on its back, or does not stop rolling	0 1 2 3

Sweating	No signs of sweating Warm or damp to touch, no sweat or wet spots visible Wet spots visible, no droplets or streams Excessive sweating, may include streams or droplets	0 1 2 3
Tail flicking (do not count flicking to chase off insects)	No tail flicking Occasional tail flicking (1-2 times/5 min) and/or holds tail away from body Frequent tail flicking (3-4 times/5 min), may hold tail away from body Excessive tail flicking (>5 times/5 min)	0 1 2 3
Kicking at abdomen (= with hind leg)	Quietly standing, no kicking Occasional kicking at abdomen (1-2 times/5 min) Frequent kicking at abdomen (3-4 times/5 min) Excessive kicking at abdomen (>5 times/5 min)	0 1 2 3
Pawing at floor (number of episodes)	Quietly standing, does not paw at floor Occasional pawing at floor (1-2 times/5 min) Frequent pawing at floor (3-4 times/5 min) Excessive pawing at floor (>5 times/5 min)	0 1 2 3
Head movements	No fast movements, head mostly at same height/in same direction Occasional head movements laterally or vertically, looking at flank (1-2 times/5 min) Frequent and fast head movements laterally or vertically, looking at flank (3-4 times/5 min) Excessive head movements, excessive looking at flank (>5 times/5 min), biting at flank (>1 time/5 min)	0 1 2 3
Pain sounds	No audible signs of pain Occasional teeth grinding or moaning (1-2 times/5 min) Frequent teeth grinding or moaning (3-4 times/5 min) Excessive teeth grinding or moaning (>5 times/5 min)	0 1 2 3
Overall appearance, reaction to observer(s)	Quiet but alert, approaches/turns to observer Alert, no reluctance to move, obvious reaction to sounds and/or movements Restless, constantly moving, exaggerated reaction to sounds and/or movements Stupor: the horse is not moving, head is lowered, reluctance to move	0 1 2 3
Reaction to palpation of painful area	No reaction to palpation Mild reaction to palpation Resistance to palpation Violent reaction to (attempt to) palpation	0 1 2 3
Total		.../42

Appendix C: Facial Expression Pain Scale used simultaneously in this study

Date:
Time:
Observer:

Observation:
0 hours
~ 1 hour if given NSAIDs at clinic
2 – 3 hours if pain management is unknown or was administered before arrival at clinic → optional
first morning
second morning

Head	Normal head movement/interested in environment Decreased movement No movement	0 1 2
Eyelids	Opened, sclera can be seen in case of eye/head movement Eyes opened further or tightening of eyelids. An edge of the sclera can be seen for 50% of the time Obviously more opened eyes or obvious tightening of eyelids. Sclera can be seen more than 50% of the time	0 1 2
Focus	Focused on environment Less focused on environment Not focusing on environment	0 1 2
Nostrils	Relaxed Slightly widened Obviously widened, nostril flaring and possible snorting/audible breathing	0 1 2
Lips/mouth corners	Relaxed Slightly raised and/or tense Obviously raised and/or tense	0 1 2
Muscle tone	No fasciculation Mild fasciculation Obvious fasciculation	0 1 2
Flehmen and/or yawning	Not seen Seen	0 2
Teeth grinding and/or moaning	Not heard Heard	0 2
Ears	Position: directed towards sound/clear response with both ears or ear closest to source Delayed /reduced response to sound Position: facing backwards/no response to sound	0 1 2
Total		.../18

Appendix D: Temperament testing concept (based on Hausberger et al. 2009)

Test	How	Score categories	What is tested
Open Field /novel environment	VIDEO – First 2 minutes without observers in stable	OF	fearfulness
Sound			
Positive	Drop food pellets in a bucket (out of sight)	Sound	attentiveness
Negative	Loud noise (whistle) out of sight	Sound	fearfulness
Human approach			
Acute	VIDEO – First reaction to clinician, assistant or observer entering stable	HA	reactivity
Test	Horse in stable: wait until horse stops paying attention to people and is not eating. Then stand directly in front of stable, look at the horse – note reaction in the first 5 seconds.	HA	reactivity
Novel Object			
Rubix cube	Stand 2 meters in front of horse and offer the object (see notes below). Watch reaction for 5 seconds.	NO	attentiveness
Touch sensitivity			
Palpation of abdomen	VIDEO – initial reaction to palpation of abdomen either by observer or clinicians	touch	sensitivity
Skin sensitivity	Lightly trace a pen or finger along the horse's flank from shoulder to loins.	touch	sensitivity

OF: classification **Open Field**

1. Exploratory: horse calmly enters stall; smells bedding and walls; walks around
2. Neutral: horse stands still
3. Evasive: horse stays alongside the walls and does not stand in the middle of the stall

Sound: classification **Sound Tests**

1. Interested: horse raises head; looks in direction of sound; may move head in direction of sound
2. Neutral: does not appear to react, may move one ear
3. Fear: raises head; pulls/jumps away from noise; acute reaction

HA: classification **Human Approach** (*Welzijnsmonitor Paardenhouderij*)

1. Interested: horse moves toward person in a friendly manner; looks up or around; stretches neck toward person; smells person or touches them
2. Neutral: horse does not move in the direction of person; only turns ears toward person without turning the head
3. Threaten: horse does not clearly move toward or away from person, but does not appear friendly (flattens ears, bares teeth, threatens to kick etc.)
4. Aggressive: horse makes a clearly threatening and/or aggressive movement toward the person
5. Evasion: horse makes a clearly startled movement away from person

NO offering: **Rubix cube Novel Object** (*Welzijnsmonitor Paardenhouderij*)

1. Approach horse from the front in a calm manner: stop at 2 meters in front of horse with cube in hand behind the back and click tongue twice.
2. Stretch arm with cube forward, with the back of the hand facing upwards (cube is not visible to the horse).
3. Turn the hand so that the cube becomes visible.
4. Observe the (main) reaction of the horse for 5 seconds.

NO: classification **Rubix cube**

1. Touching: the horse immediately stretches its neck and possibly takes a step forward and touches the cube within 5 seconds.
2. Reaching: the horse moves forward, carefully stretches the neck or may take a step forward; can smell the cube but does not touch or does not touch within 5 seconds.
3. Neutral: the horse does not stretch its neck and does not take a step forward.
4. Turn away: the horse turns (head and/or entire body) calmly away from the cube.
5. Evasion: the horse flinches, steps back/is scared of the cube.

Touch sensitivity: classification **Touch** (adapted from *Welzijnsmonitor Paardenhouderij*)

1. No indication of tension and/or sensitivity (light tensing followed by relaxation of back muscles at touch is considered normal).
2. Indication of heightened sensitivity: slight dodging/ “hard” back / tensing; *no* or mild behavioral reaction (such as turning ears backward).
3. Indication of (very) high sensitivity: mild to severe reaction to pressure; pulling away from the touch, back flinching, kicking or an attempt to, biting, turning away from the researcher, walking away