Author/s:
Anderson, Katrina

Title:
Use of behavioural signs for riskbased prevention of catastrophic breakdown in racehorses

Date:
2020

Persistent Link:
https://hdl.handle.net/11343/242021

Use of behavioural signs for risk-based prevention of catastrophic breakdown in racehorses

by

Katrina Anderson

0000-0001-7249-5629

Master of Veterinary Science
May 2020
Faculty of Veterinary and Agricultural Sciences
University of Melbourne

Submitted in fulfilment for the requirement for the degree of
Master of Veterinary Science

Abstract:

Bone and joint injuries in Thoroughbred racehorses typically originate in areas of intense loading. Despite an increasing understanding regarding the pathogenesis of bone fatigue, early detection prior to fatigue fracture is difficult. This thesis focuses on understanding equine behaviour and its potential to reflect orthopaedic pain to assist in the early identification of horses at risk of injury. Previous research suggests an association between facial expression and the presence of equine lameness. In cattle, motion sensors have been validated to record recumbency behaviour that indicates disease without the need for continuous observation. This thesis aimed to investigate monitoring of behavioural patterns as predictors of injury in racehorses.

In the first study, Thoroughbred racehorses were monitored at rest with motion sensors placed on forelimbs. Sensor data was collected, and an algorithm developed to identify certain behaviours. Algorithms were validated through video observation and calculation of algorithm sensitivity and specificity. There was excellent agreement for standing, lying (>99%), and for stepping (>86%) behaviour patterns suggesting potential for long term, objective, remote monitoring of horses in the stable environments.

In the second study, previously validated facial grimace scales were used to score pain from photographs of Thoroughbreds, captured from video recordings during trot up examinations. Cohen's kappa (κ) was used to measure the proportion of agreement between raters. The kappa was ordinal weighted for behaviours scored into more than two categories. Mean inter observer agreement was moderate (κ 0.45; 95% CI 0.36, 0.55), and the sum of facial scores showed no difference between lame and horses whose asymmetry fell below thresholds considered ‘fit to race’ measured by inertial sensors. Moderate mouth strain (HGS) and tense and extended upper lip (FEReq) were associated with gaits that were less asymmetric ($p=0.043$ and $p=0.027$, respectively). Exposed sclera was associated with lameness ($p=0.045$) in hand, at the trot. There were few associations between lameness status and facial predictors, limiting the potential use of facial expression for prediction of orthopaedic pain during pre-race lameness examinations.

In conclusion, motion sensors are sensitive to detect horse movements and have potential use in longitudinal monitoring of horse behaviour. Future research in this area includes prediction of the use of behavioural indicators to detect musculoskeletal pain in racehorses that could contribute to the development of a risk-based injury prevention model. The facial grimace scales had limited ability to differentiate lame from horses considered ‘fit to race’ in this cohort of horses. External factors associated with racehorse management,

including environmental stress, may be one reason facial grimace scales are not effective for the recognition of subtle lameness in this cohort.

Declaration of originality

This thesis contains no material accepted for a degree by the University or any other institution, except information appropriately acknowledged in the thesis. I declare that all material was not written by any other person except where due acknowledgement is made. The thesis is fewer than the maximum word limit in length.

Signed: ……………………………………………………………… Date: …………………………………………………….

Statement of Authorship

This thesis includes papers for which Katrina Anderson (KA) was not the sole author. K Anderson was the principal researcher as she completed data collection, analysed the data and wrote the manuscripts. However, she was assisted by the co-authors whose contributions are acknowledged below:

1. Validation of accelerometers to detect and predict horse behaviour (Unpublished material not submitted for application)

Katrina Anderson, Peta L Hitchens, Elizabeth Walmsley, Andrew Fisher and R Christopher Whitton

K Anderson was responsible for obtaining approvals, design and management of the study, data collection and analysis, development of the algorithm, and compiled the initial draft of the manuscript.

P.L Hitchens, C Whitton, A Fisher and E Walmsley worked on concept and design of study, as well as interpretation of data analysis and revised the manuscript.

2. Lack of correlation between facial grimace scales and gait asymmetry in TB racehorses at the trot (Unpublished material not submitted for application)

Katrina Anderson, Peta L Hitchens, Elizabeth Walmsley, Ashleigh Morrice-West, Andrew Fisher and R Chris Whitton,

K Anderson was responsible for obtaining approvals, design and management of the study, data collection, data management and analysis, and compiled the initial draft of the manuscript.

P.L Hitchens, C Whitton, A Fisher and E Walmsley assisted with study design, and data analyses and interpretation of data, and revised the manuscript. C Whitton helped with data collection.

A Morrice-West helped with data collection, data analyses and interpretation.

This research was funded by Racing Victoria Ltd. (RVL), the Victorian Racing Industry Fund (VRIF) of the Victorian State Government, and the University of Melbourne. Katrina Anderson is funded by a Melbourne Research Scholarship.

Signed by first named supervisor, Prof R Chris Whitton:

Signed: ……………………………………………………………… Date: ……………………………………………………….

Acknowledgements:

There are so many wonderful people I would like to thank for their support over the length of this thesis.

First and foremost, a huge thank you to all my supervisors: Prof. Chris Whitton, Dr. Peta Hitchens, Elizabeth Walmsley, and Prof. Andrew Fisher, for their thought provoking comments and encouragement, but also for their motivation to continually push me to reach further in my research. A special mention to Dr. Adelene Wong for her important contribution to organising and scheduling every milestone.

I feel extremely grateful and fortunate to have worked under the guidance of Prof. Chris Whitton. I would like to express my gratitude for his continuous support of my research, for his patience, and immense knowledge.

I have thoroughly enjoyed working with Dr. Peta Hitchens, her door was always open (metaphorically, as she didn't have a door) whenever I had a question about my research, writing or most probably statistics. We always enjoyed a laugh about seemingly ridiculous situations.

I would like to express my deepest gratitude to Liz Walmsley for her ongoing advice and support, not only research related. I will always appreciate our daily doggy playdates and endless chats about our wonderful dogs, amongst discussions related to my project.

I would also like to acknowledge Prof Andrew Fisher, the Director of the Animal Welfare Science Centre at the University of Melbourne, I am grateful for his very valuable comments on this thesis.

I thank my fellow research student, Ashleigh, who was an absolute star throughout this thesis, not only for the stimulating discussions and help with statistics, but also for her friendship. Thank you for making our daily walk to get a cup of tea seem like such an adventure.

Endless appreciation to all the vets, nurses and staff whose assistance was a milestone in the completion of this project.

This research was made possible by the financial support of the University of Melbourne and funding from Racing Victoria. Additional appreciation to Denny and Logemas for the loan of the motion sensors.

The team at Equinosis (Lameness Locator) for their support throughout, promptly answering any questions I had and always willing to help when needed.

Susanne Ellens gave me invaluable help with developing the algorithm, with her guidance I was able to successfully complete this project.

I would also like to thank volunteers, Zoe and Leonie, for their time in the pain scoring study.

My research would not have been possible without the willingness and support of Ciaron Maher and David Eustace, and their staff, who allowed me to conduct all data collection at their racing stable without any hesitation.

I must express my very profound gratitude to my boyfriend Matt, my running friends, and my flatmates, for their unwavering support and continuous encouragement throughout the two years of study. Thank you for celebrating my successes, listening to my ideas, problems and passions. Thank you for motivating me through the ups and downs involved with researching and writing a thesis. This accomplishment would not have been possible without them.

Mum and Dad, you taught me every value I needed to succeed (you are also the reason I am passionate about horses and horse racing). Making you proud has always been the biggest motivation throughout. You are always the reason I never give up, love you.

And finally, a special thank you to my four-legged friend who spent long days at the office with me, providing constant emotional support. Thanks Roo.

Table of Contents

List of Tables:

List of Figures:

List of Abbreviations:

AAEP	American Association of Equine Practitioners
BMD	Bone Mineral Density
BV/TV	Bone Volume Fraction
CMI	Catastrophic Musculoskeletal Injury
CPS	Compositive Pain Scale
CT	Computed tomography
FACS	Facial Action Coding System
FEReq	Facial Expression in Ridden horses
HDmax	Maximum displacement of pelvis during trot
HDmin	Minimum displacement of pelvis during trot
HGS	Horse Grimace Scale
HR-pQCT	High-resolution peripheral quantitative CT
IMU	Inertial Measurement Units
IRU	Increased Radiopharmaceutical Uptake
LL	Lameness Locator
MSI	Musculoskeletal injury
MRI	Magnetic Resonance Imaging
NRS	Numerical Rating Scale
OA	Osteoarthritis
Pmax	Maximum displacement of pelvis during trot
Pmin	Minimum displacement of pelvis during trot
POD	Palmar/plantar Osteochondral Disease
TB	Thoroughbred
VAS	Visual Analogue Scale

Chapter 1: Introduction

1.1 Background summary

Musculoskeletal injuries are the most common cause of lost training days and wastage in racehorses, resulting in substantial welfare and economic implications [1, 2]. Between 2016/2017 there were 35,309 Thoroughbred (TB) racehorses competing in races nationwide, which contributed $671,161,297 AUD to the Australian economy [3]. In Victoria between 1989 and 2004, the risk of TB racehorse fatality was 0.44 per 1000 race starts, lower than the global incidence [4, 5], with 73% of fatalities due to catastrophic musculoskeletal injury (CMI) (0.32 per 1000 starts) [6].

Post-mortem examinations have confirmed that most catastrophic fractures are associated with pre-existing stress fractures or pathologic conditions [5, 7-9]. Many skeletal injuries in racehorses develop gradually as the integrity of the bone is overcome by repetitive loading during exercise. Fatigue fractures are most common in human athletes, ballet dancers, military recruits and racehorses [10]. Fatigue injury occurs when microdamage – a product of cyclic loading – accumulates faster than it can be repaired [11]. Despite advancements in diagnostic imaging and knowledge of pathophysiology of bone fatigue, detection of early signs of fatigue to identify horses at high risk remains difficult. This is because clinical signs are not always present, and when they do become apparent it is often too late. The use of imaging equipment provides the potential for early diagnosis, but finances and practicality may limit the feasibility of widespread use of advanced imaging as a screening tool.

Monitoring behavioural changes may provide an additional method for early detection of pain preceding injury. Behavioural changes are observed in horses in association with the presence of post-operative musculoskeletal pain [12, 13], visceral pain [14, 15] and with lameness [16, 17]. Further research is required to determine whether correlations exist between behaviour, lameness and musculoskeletal injuries.

1.1 Bone fatigue

Racehorse limbs are subjected to repetitive loading during training and racing. Without opportunity for the bone to rest and repair, microcracks may coalescence into a larger, catastrophic fracture [18]. Degradation of material properties of bone is closely linked to the fatigue life, defined as “the number of cycles an applied load can be sustained by a structure before catastrophic failure occurs” [19]. The fatigue life of bone can be depicted using an S-N curve showing a linear relationship between stress and strain; where the number of cycles to failure decreases as the size of load per area increases (Figure 1:1). Lower magnitudes of load cause less microdamage and therefore more cycles of low load can be endured before failure as compared with high magnitude loading. High stress on the bone, i.e. galloping, results in a lower number of cycles to failure [20]. Verheyen *et al*. found horses that exceeded 7700 cycles at canter and 880 cycles at gallop were at a higher risk of injury [21]. The accumulation of pathology leading to fatigue injury is supported by the increased risk of fracture in older horses because they typically have had longer careers [22].

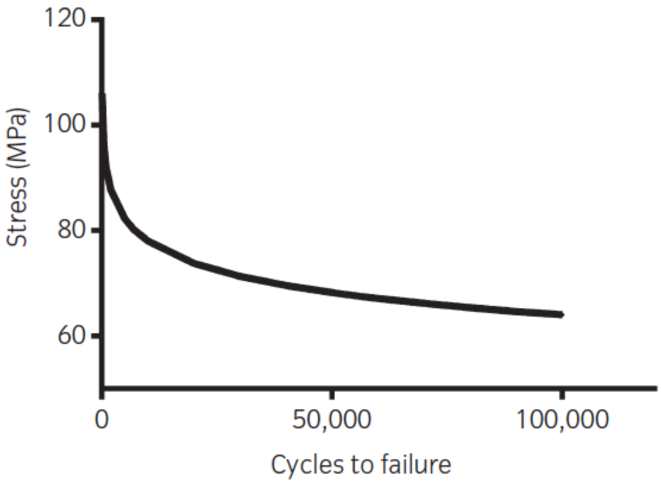


Figure 1:1 S-N graph: Illustrates the exponential relationship between stress and strain. Adapted from Martig et al. (2014)

1.1.1 Bone adaptation to load

Microdamage stimulates biological processes that alter the biomechanical properties of bone. This process is explained by Wolff's law, that bone changes its external conformation and internal architecture in response to the stresses it is exposed to [23]. These responses to mechanical forces can be explained by theoretical mathematical laws [24]. Bone is a dynamic tissue that can adapt to changes in load [25] through modelling and remodeling processes, resulting in changes to bone structure [19] and replacement of damaged bone with new bone [26, 27]. To achieve optimal function and increase resistance to failure, bone adaptations through modeling, include filling of trabecular spacing, increasing bone volume, increased cortical thickness, and changing shape [19, 28, 29]. Remodeling actively occurs during rest periods, but is reduced during high intensity training, thus allowing damage to accumulate [11]. Inadequate rest periods reduce the opportunity for the bone to repair [11, 19, 30-32]. On the contrary, the initial phase of bone remodeling may be detrimental to bone strength due to the temporary increase in porosity [33]. The increased ratio of pores in a given volume, weakens bone and has the potential to accelerate fatigue [18].

1.2 Early detection

1.2.1 Pathology of bone fatigue

Post mortem examinations of horses with fatal fractures have often confirmed the presence of pre-existing lesions in both affected and contralateral limbs [5, 7, 34]. Under repetitive load, changes to mechanical properties of bone can be observed. The point at which lesions become significant for predicting a horse at risk of fracture have not been determined, and moreover, there is no one set rate at which lesions develop. Fatigue injury to joint surfaces can manifest in a number of different clinical syndromes such as subchondral bone injury, osteochondral fragmentation, articular cartilage degradation, palmar osteochondral disease and intra-articular fracture [7, 18, 35, 36]. Palmar osteochondral disease (POD) is a common finding in the third metacarpal and metatarsal bones of Thoroughbred racehorses [37, 38]. Although POD is associated with fewer starts and overall poorer performance, the high incidence suggests horse can continue to train despite its presence [31, 35, 39]. POD may be a protective against CMI in this location [40].

Sclerosis, or increased bone density, is considered the natural response to cyclic load but is also a response to injury [41]. For example sclerosis may co-exist with localised microdamage and is identified commonly in racehorses with fracture [42]. Increased bone volume fraction (BV/TV) was observed in the distal metacarpus of horses with fatal limb injury, and can help distinguish sesamoid bones of fractured horses from controls [43, 44]. In another study, BV/TV had only moderate sensitivity and specificity to predict injury [44]. Depth of dense subchondral bone has merit for predicting fracture; for every millimetre increased subchondral bone depth, the likelihood of fracture increased fourfold [45].

1.2.2 Diagnostic imaging techniques

Diagnostic imaging has the potential to identify some of these pathological responses, but it is unclear what extent of pathology should be considered high risk for future CMI. To be useful as a screening tool an imaging modality requires various attributes including high resolution, cost efficiency, sensitivity for diagnosis, and be minimally disruptive to racehorse training.

Radiography cannot detect a decrease in mineralisation of less than approximately 30%, and therefore changes in bone mineralisation may be undetectable in the early stages of disease, potentially delaying identification of impending fracture [46, 47]. The bone response can be observed on radiographic imaging, by identifying periosteal or endosteal abnormalities, as well as sclerosis [48]. However traditional radiography is limited to 2-dimensional images of a 3-dimensional structure. Computed tomography (CT), 3-dimensional imaging using x-rays, allows evaluation in greater detail and of structures that would otherwise appear superimposed [49]. Sensitive to density differences in tissue as small as 0.5%, CT is capable of identifying coalescing cracks absent in radiographs [43, 50-52]. The cost and need for general anaesthesia limits its clinical use with racehorses although standing CT has recently become available [51]. Standing CTs performed using small peripheral quantitative computed tomography has been described (pQCT) [53]. High-resolution peripheral quantitative CT (HR-pQCT) has been used on cadavers to describe the effect of training on gradient of bone mineral density and maybe a useful modality to evaluate bone and predict risk for fracture [54, 55]. The high resolution of micro-CT imaging can detect changes in bone morphology, however this is currently not attainable in vivo [43].

High-filed magnetic resonance imaging (MRI) has been used to measure the depth of dense subchondral/trabecular bone within the lateral parasagittal groove to provide an optimal cut off of 16mm to best differentiate horses with and without fracture [45]. Using this cut off, the sensitivity of this ‘test’ was 79.2% and the specificity 98% [45]. The opportunity to avoid general anaesthesia through use of low field MRI machines makes this an appealing option for Thoroughbred trainers. The ability to image horses whilst standing, rather than under anaesthesia ay improve uptake of the process by trainers. However, the small field of view and limited spatial resolution, limit the use of MRI as a screening tool since the region of interest must be known. Evaluation of cortical bone fracture is also limited using sMRI [56].

Nuclear scintigraphy has proven to be highly sensitive to early changes in bone metabolism by using radioactive properties to determine blood flow and osteoblastic activity [57, 58]. Scintigraphy is often used when the source of pain is difficult to localise. Lack of specificity [31] and practicality limit its use as a screening tool for pre-race examinations. Increased radiopharmaceutical uptake (IRU) in the parasagittal groove is difficult to differentiate from condylar uptake [31]. Ultrasonography is not widely used as a screening tool, but is used to diagnose fractures in areas not easily accessible to radiographic examination such as pelvis, ribs and scapula [59].

1.2.3 Biomarkers

The usefulness of serum biomarkers in Thoroughbred racehorses as a preventative measure for fracture has been investigated over the last decade [60-62]. Biomarkers indicate a biological or pathogenic process that can be objectively measured and used to detect early changes in tissue synthesis or breakdown [63]. Quantitative identification of biomarkers may provide the most sensitive method to detect changes in bone metabolism and precede clinical signs of musculoskeletal injuries [64]. Markers of bone formation and resorption vary at different stages of training [65]. Markers of bone turnover decrease during race training which correlates with the inhibition of remodeling during high speed exercise [26, 66]. A longitudinal study revealed that significant changes in biomarkers were observed based on the injury sustained, with horses correctly classified as injured 74% of the time [60]. However, significant patterns of biomarkers to predict pathology changes prior to injury is yet to be determined [63].

Biomarkers have shown the ability to identify pain and stress in horses. Heart rate, respiratory rate and blood pressure measured noninvasively were moderately sensitive for detection of pain in horses with orthopaedic problems and with colic, but did not differentiate controls from post arthroscopy cases [13, 14, 17, 67-69]. Cortisol levels were significantly elevated in horses with orthopaedic and post-operative pain [14, 67]. However, other equine studies have showed cortisol was not a good indicator of pain [17, 70].

1.3 Lameness

Equine lameness is an alteration of gait due to a structural or functional disorder of the locomotor system [71]. Lameness is the most commonly reported reason for failure to train and reduced racing performance, affecting 53% of the racehorse population at a given time [2, 72]. A survey of UK racing yards reported that lameness accounted for approximately 82% of total days lost from training in two and three-year-old racehorses, with 20-30% cases due to stress fractures [2]. Lameness is a clinical sign, commonly attributed to orthopaedic pain [71]. The detection of lameness is centred around visual detection of asymmetrical movement between the left and right side. Buchner *et al.* analysed the equine gait under induced lameness conditions and determined asymmetry of head and tuber sacrale during movement were the best indicators of lameness [73]. Horses redistribute body mass away from the affected limb resulting in asymmetric movements of the trunk [74]. The trunk velocity is reduced during the impact phase of the lame limb, with the trunk kept higher off the ground throughout the stance phase [73]. Many different components of gait have been studied and used to detect forelimb and hindlimb lameness in horses, including stride length, duration of weight-bearing on the affected and unaffected limbs, head nod, hip hike and joint angle or fetlock extension [50, 73, 75-77]. Arguably the current gold standard for detection of asymmetry is the use of objective measurement systems that describes gait using a number of motion sensors [78-80].

The "law of sides" is used to explain patterns of asymmetry that occur as a result of redistribution of bodyweight during the stance phase of the stride. This forward movement of body weight to compensate for hind limb lameness, causes a lowering of the forequarters as the contralateral forelimb is in the stance phase, an asymmetry that would be observed to mimic an ipsilateral forelimb lameness [81]. A horse observed lame in both the forelimb and hindlimb could have a true primary and a compensatory (false) asymmetry or a true primary and true secondary limb lameness. Compensatory lameness can occur as an ipsilateral lameness or a contralateral lameness. Ipsilateral lameness, defined as forelimb and hindlimb lameness on one side, may be due to a primary hindlimb lameness with a compensatory, false forelimb lameness [82]. In contralateral limb pairs a true forelimb and compensatory contralateral hindlimb asymmetry is the likely explanation.

1.3.1 Subjective Evaluation

Clinical observations by trained practitioners remain the most convenient method for assessing lameness in a horse. The varying nature of lameness makes diagnosis difficult and despite the use of standardised scales to objectively evaluate lameness, inter-observer reliability is low [76, 83]. The current accepted scale used to assess lameness is that of the American Association of Equine Practitioners (AAEP), which uses 5 grades to describe severity; where zero is no lameness observed under any circumstances and five is non-weightbearing lame. Studies showed when using the grade system AAEP, veterinarians agreed in only 62% of cases whether a limb was lame or not [84] [85]. Interobserver agreement is typically low irrespective of scale used [76, 83], most likely due to the lack of sensitivity of scales and subjective nature of visual examination.

1.3.2 Objective Lameness Examination:

Use of objective measures of gait symmetry seek to address these issues through measurement of displacement of fixed points on skeleton [80, 86]. Quantitative techniques such as kinematic analysis and kinetic analysis have been developed to assist veterinarians.

1.3.2.1 Kinematic Methods:

Kinematic methods analyse movement of internal and external body segments during locomotion [85]. Three dimensional optical motion capture (OMC) uses cameras to provide three-dimensional data while the horse is moving over ground or on a treadmill [87]. Using camera and body markers, the trajectory and motion of joints are tracked to compare between right and left [88]. The need for many OMC cameras and supporting infrastructure limits its use in a clinical setting [89]. Therefore, alternative methods using inertial measurement units (IMUs), mounted to different anatomical locations on the horse have allowed for over the ground locomotion analysis [90]. These devices can be attached at key locations on the body and measure acceleration and angular velocity using accelerometers and gyroscopes. Agreement between IMU and OMC has been established [85]. The ability to objectively measure horse movement and quantitatively assess lameness has seen wide acceptance within the equine veterinary field [80, 84, 91-93].

1.3.2.2 Kinetic Methods:

Kinetic methods analyse how a body moves by objectively measuring the forces generated through movement [85]. Unloading of the affected limb will result in measurable decreased ground reaction force (GRF) during the stance phase, with peak forces decreased with increased severity of lameness [94, 95]. Detection of GRF changes appears more sensitive than observation for lameness [95]. Force plates are considered the gold standard however controlled conditions and time required reduces clinical applicability [89]. To address these limitations, technological advancements to measure GRF developed a device placed between hoof and shoe [96] and a force measuring treadmill [75]. However, none of these methods are easily incorporated for daily use outside a scientific environment.

1.3.3 Lameness Locator:

A commonly used inertial sensor system – developed to detect asymmetry of gait – is the Lameness Locator® (Equinosis, St Louis, Missouri, USA), which uses two accelerometers attached to the head, and one gyroscope attached to the right forelimb [80]. The examined horse is trotted over ground, while the attached sensors collect data that can be wirelessly transmitted to generate an immediate gait analysis of stride by stride differences. The maximum and minimum head (HDmax and HDmin) and pelvic (PDmx and PDmin) height differences are calculated for each stride, which is reported as a mean difference (in sensor displacement) over each trial. For displacement of minimum amplitude, the sign refers to the side of asymmetry, with positive measurements indicating right sided lameness and negative indicating left lameness. Displacement of maximum amplitude refer to measurement of the upward position following stance phase. Due to the changes in horse gait, both the push off and the impact phase affect the maximum displacement of head and pelvis. The limb and phase of stride affected can be determined using the sign of displacement for both the minimum and maximum amplitudes.

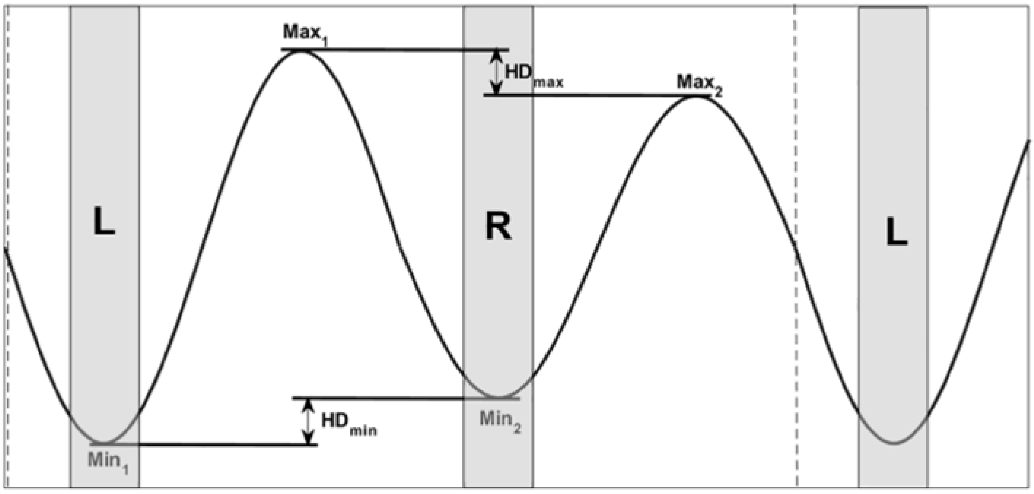


Figure 1:2: Example of vertical head movement in a horse with right forelimb lameness resulting in positive HDmin and HDmax values (adapted from Rhodin et al.)

Measured displacement considered to be in the normal range is the threshold for determining a horse lame. Keegan et al. determined thresholds of >6mm for head displacement and >3mm for pelvis displacement, based percentages of amplitude derived from 136 horses evaluated for lameness, poor performance and pre purchase examination [80]. Artificially produced asymmetry showed that 25% asymmetry was required to be detected by the human eye regardless of experience [97]. The higher sampling frequency (200Hz) of the LL compared to the unaided human eye (15-20Hz) allows detection of subtle lameness and has proved repeatable for use in a clinical setting [80, 84]. Interpretation of the significance of a measured displacement requires evaluation by a trained clinician in addition to a full lameness examination including straight trot up, lunging, flexion tests and diagnostic analgesia [79] [98].

Horses can perform despite gait asymmetries. There have been numerous studies demonstrating a high incidence of horses perceived sound by their owners that had asymmetries in gait [79, 99, 100]. Forty-seven percent of sport horses with pain-related gait abnormalities had not been either identified or acknowledged by trainers or riders [101]. Pfau *et al.* found 75% of racehorses that were thought fit to race fell outside predetermined thresholds [102]. Higher thresholds (>14.4mm for HDmin, and >7.5mm for PDmin) have been proposed as a more suitable barometer of "fitness to race" based on pre-race evaluation of Thoroughbred racehorses by a group of experienced racetrack veterinarians [78]. Although to establish thresholds, agreement between pairs of observers was only poor to fair. Overall agreement was 0.60. Maximum displacement asymmetries lacked specificity, which confirms the difficulty of veterinarian to agree on which limb was lame [78]. Maximum amplitudes of head and pelvis were excluded from thresholds.

1.4 Pain

Pain is defined as an unpleasant sensory and emotional experience associated with actual or potential tissue damage [103]. In the presence of harmful stimulus pain, nociceptors convey electrical signals to the thalamus and sensory cortex of the brain, where it is consciously perceived as pain [104]. The purpose of nociception is to protect the individual from present or future harm through physiological and behavioural changes [104]. Common behavioural changes include: increased aggression [105], reduced human interaction [14, 67], changes to posture, and lameness [67].

1.4.1 Behaviour:

As horses lack the ability to verbalise how they feel, behaviour can provides useful insight into the wellbeing and help recognise presence of pain [106]. Behaviours associated with pain are often considered non-specific, however, observed collectively they may be useful to evaluate clinical signs of underlying pathology [69, 107]. Pain detection and quantification in horses is dependent on subjective observation [107]. Standardized pain scales account for frequency, duration and intensity of abnormal behaviour [67, 107]. Multiple techniques are used in practice. The visual analogue scales (VAS) are sensitive but limited by inter-observer reliability in equine studies [68]. The numerical rating scale (NRS), is a discontinuous scale numbered from zero to ten which was more repeatable than the VAS [14], but less sensitive to small changes. Simple descriptive scale (SDS) uses descriptions to interpret pain, and composite pain scales (CPS) combine both behaviour and physiological parameters [67].

Behavioural changes resulting from induced orthopaedic pain in horses have been described [67, 68, 108], but behaviour can be nonspecific therefore a combination of findings is needed to assess orthopaedic pain in horses [17]. Orthopaedic pain subsequent to amphotericin-B injection in the tarsocrural joint, suggested the most significant indicators for pain were posture and response to palpation [67]. However, the nature of this pain induced an acute inflammatory response which may be different to pain as a result of accumulation of bone damage as occurs in fatigue processes. For visceral postoperative pain, pawing the ground, head movements and inactive behaviour were most prevalent [109]. Following gastrointestinal surgery, horses spent more time inactive when compared to control groups which was perceived as pain [14]. Unusual stance and weight shifting are specific pain-related behaviour associated with decreased weightbearing, therefore relevant to an investigation of behavioural changes associated with orthopaedic pain [69]. Weight-shifting has proved reliable for detecting discomfort in horses suffering from laminitis. Rietmann *et al.* manually recorded frequency of weight shifting from video recordings and established that frequency of weight shifts was positively correlated with an increasing severity of the disease [17]. Following administration of nonsteroidal anti-inflammatory drugs (NSAIDs) a decrease in weight-shifting frequency was observed, correlating the observed behaviour with the presence of discomfort [17].

1.4.2 Horse Grimace Scales

Facial expressions are used to assess pain in people due to their dynamic nature, changing rapidly in response to stimuli and internal states, which cannot be suppressed [110]. The Facial Action Coding system (FACs) objectively describes facial expression, incorporating underlying facial musculature and muscle movement relating to the eyes, nose, cheeks and mouth [111, 112]. Originally designed for humans, the FACs has been modified for animals through comparisons in facial anatomy. Grimace scales using FACs have been successfully utilised in rodents, rabbits, and horses [12, 113-115].

The Horse Grimace Scale (HGS) incorporates six facial units associated with pain: ear position, orbital tightening, eye area tension, strained chewing muscles, strained nostrils, and strained mouth [12]. Each facial unit is scored using a 3-point scale (0 = not present, 1= moderately present, 2= obviously present) and the sum of these six scores will yield the HGS score. The scale was validated in a blinded study using photographs of 40 horses following castration surgery with high inter-observer reliability (0.92) [12]. Overall HGS scores were higher in horse's post-surgery compared to pre-surgery and the control group. Additionally, an overall pain score demonstrated 73% accuracy to differentiate post-operative horses. Dalla Costa *et al.* further investigated the HGS using video observations of laminitis patients and determined no significant difference between scores from images and videos [116].

The facial expressions in ridden horses (FEReq) protocol was developed to improve pain recognition in ridden horses [16, 117]. The FEReq could be assessed relatively consistently by a small group of individual assessors [117] and pain scores were higher for lame horses than nonlame horses (P<0.001) [16]. A pain score was allocated to each of three parameters using a weighted system based on the occurrence of the behaviour observed in lame horses during development. Mean values for facial pain scores were higher in lame horses than sound horses, 8.7 and 6.1, respectively [16]. Significant facial expressions included: twisted nose, eyes partially closed, intense stare, and tension caudal to the eye [16].

Dyson et al. developed a new ethogram using video analysis to detect dynamic movements and duration of behaviour, associated with the presence of musculoskeletal pain [101]. Diagnostic analgesia was used to validate behaviours as being associated with pain, and found certain behaviours were more strongly correlated with lameness than others. Fifteen behavioural markers occurred significantly more often in lame horses and another four markers were only observed in lame horses [118]. The average scores were 6.2 times higher for lame horses compared to the same horse after the administration of analgesia to reduce lameness.

It is unclear whether video or photographs are more sensitive for grading severity of pain, nor is it established what method is more applicable for use in routine assessment or a clinical setting. Interpretation of pain as determined from a single frame may not reflect actual pain grimace scores. Side and front on images had large variation in HGS reliability for different parameters, reliability of “wrinkles between the nostrils” was 0.57 when using side images, compared to 0.76 using front on images [12, 116]. In ridden ethograms, partial or complete eye closure was validated by photographs [12, 117], but was not observed using video observation [101].

1.4.3 Quantification of pain

Current methods to quantify horse behaviour are time and labour intensive [13]. Pritchett *et al.* reduced manual intensive monitoring by calculating the duration of behaviour at designated time increments and then extrapolating them to estimate the proportions of each behaviour over a period [14]. Although time efficient, this method is less sensitive. Horse trainers have responsibility to maintain horse wellbeing through visual monitoring, however, they can only observe behaviour displayed in the present moment. Moreover, recent research has shown horses in pain hide their discomfort in the presence of people [119]. Over the past decade, the racing industry has reported a 16% increase in the total number of racehorses, while there has been a 32% decline in the number of trainers, meaning larger training yards and greater challenge to observation of behaviours one on one [3, 120]. Practical methods of quantifying horse behaviour during the absence of personnel are required. The implementation of a device capable of monitoring activity would be a highly desirable addition to the routine assessment of racehorses in training, with the potential to improve animal welfare.

1.4.4 Accelerometers:

Over the past decade methods of recording physical activity has changed following the application of inertial measurement units (IMUs) that automatically record movements [121]. An IMU is a type of sensor that measures acceleration, angular velocity and sometimes magnetic field. They are composed of a 3-axis accelerometer and a 3-axis gyroscope (and 3-axis magnetometer). In various species IMUs have provided useful to measure lameness [122]; the activity of zoo animals [123]; and help predict upcoming birth through changes in lying behaviour [124]. Accelerometers have potential to improve quantification of animal behaviour and replace time consuming visual observation methods of animals [125].

Raw acceleration data collected from IMUs can be used to train machine learning algorithms. An algorithm is developed to identify patterns in the waveforms that correspond to a behaviour or movement [126]. To develop an algorithm the researcher must initially label behaviour manually using time-matched observation material [127]. Supervised machine learning is a computational learning method used to make predictions about unknown data. The supervised learning algorithm takes labelled data from a known (training) dataset to learn the classification model [128]. Performance of machine learning is determined by ability to accurately predict unseen values that are not part of the training set. The most common supervised classification techniques are, k-nearest neighbours, decision trees, Bayesian networks and support vector machines (SVM). SVM classifiers, ideally used for binary classification [129], have been used to predict behaviour in dairy cattle with reasonable predictive ability [130]. They have also been validated to classify eight behavioural states in dogs on a second by second basis with excellent predictive values [131].

1.5 Conclusion

Lameness is prevalent in TB racehorses and identifying features of lameness that indicate the underlying pathology that might progress to more serious injury is difficult. The objectives of this thesis were to validate a method using IMUs to identify behavioural indicators that may relate to discomfort or pain in TBs, and to identify facial features that may reflect lameness in TBs in training. To validate potential early indicators for bone fatigue, the research project investigated; (1) whether labour intensive methods of assessing behaviour can be replaced by automated monitoring systems (IMU) and; (2) whether lameness is correlated with pain-related behaviour using the Horse Grimace Scale (HGS). This information could be used in future research to address the hypothesis that bone fatigue injuries – the consequence of accumulation of microdamage – are associated with changes in behaviour, increased levels of stress, subtle lameness and increased bone turnover.

1.6 Research aims and objectives

1.6.1 General aim

The primary aim of the investigations conveyed in this thesis was to describe horse behaviour and to investigate better methods of monitoring behaviour for potential use in longitudinal studies.

1.6.2 Specific Objectives

The specific objectives of this research were:

1. To investigate the sensitivity of IMUs to detect behavioural changes.
2. Develop an algorithm to classify horse behaviour; specifically: standing, stepping, recumbency, pawing, and weightshifting.
3. Assess the inter-observer reliability of the horse grimace scale (HGS) in racehorses undergoing a lameness examination.
4. Determine whether there are significant correlations between facial expressions and lameness in Thoroughbreds.
5. To identify areas of future research in this field.

1.7 Thesis outline

The structure of this thesis is described below:

Chapter 1: Introduction

This chapter provides the background for the investigations and outlines the objectives and structure for the thesis. A literature review is provided that describes the development of catastrophic breakdown and the difficulties in detecting subtle signs of lameness, potentially associated with the onset of underlying injury.

Chapter 2: Validation of motion sensors to detect and classify horse behaviour

This chapter investigates whether labour intensive methods of monitoring behaviour (i.e. observation) can be replaced by objective monitoring (i.e. IMUs). This study aims to determine whether motions sensors are sensitive to detect subtle movements and classify horse behaviour. The contents of this chapter are intended to be submitted for publication in a peer reviewed journal.

Chapter 3: Correlation between horse grimace scales (HGS) and lameness in trot up examination

This chapter investigates the sensitivity and reliability of facial expressions to predict subtle lameness in TB racehorses. The contents of this chapter are intended to be submitted for publication in a peer reviewed journal.

Chapter 4: Discussion

This chapter summaries the important findings, implications and conclusions from this research and proposals for future research.

Chapter 2: Validation of motion sensors to detect and predict horse behaviour

2.1 Preface

In this chapter, the sensitivity of motion sensors to detect and classify horse behaviour, such as posture (i.e. frequent weight shifting of the forelimbs), recumbency duration and frequency, and movement (step frequency), are investigated. The data to develop an algorithm to classify these behaviours was collected from stabled horses. The text that follows has not been submitted for publication.

2.2 Introduction

Musculoskeletal injuries are the most common cause of lost training days and wastage in racehorses resulting in substantial economic impact to both owners and trainers [1, 2]. Two and three year old horses reportedly fail to train for approximately 21-27% of total training days [2]. Of those days lost from training, fracture accounted for 17-25% [2]. Most injuries in racehorses are due to tissue fatigue, developing gradually as bone or tendon integrity is overcome by repetitive load [11]. For bone, this is evidenced by post-mortem examinations revealing pre-existing pathology for fracture including periosteal callus formation, focal bone resorption, and microdamage [5, 7, 18]. Because most injuries develop gradually due to accumulated damage there is an opportunity for early detection [5, 7]. The use of diagnostic imaging techniques to detect injury early has been investigated, but low specificity, high cost and logistics limit their use for screening large numbers of cases [44, 45, 51, 132].

Orthopaedic pain, commonly attributed to lameness in horses, is associated with changes in behaviour including weight shifting [17], pawing the ground [67] and restlessness [13, 14]. The development of an objective, preferably real-time, system to monitor horse behaviour may enable the early detection of bone injuries prior to more overt clinical signs developing. Numerous measures to evaluate pain have been developed to score severity based on numbered or simple descriptive scales, including composite multifactorial scales and facial expression [12-14, 16, 67, 68, 76, 108]. Horses in pain are shown to spend more time inactive [14] and those specifically with orthopaedic pain have been observed unloading one limb, pawing and weight shifting [13, 14, 17, 67]. Current methods of intermittent monitoring may not accurately record frequency of pain-related behaviours [14]. Additionally, a trend towards increased racing stable size (number of horses per trainer) over the past decade makes close monitoring of horses more challenging [3, 120]. Continuous observation is not feasible nor practical for most racing stables, and horses tend to hide their discomfort in the presence of observers [119]. Therefore, an automated, objective monitoring system to aid in the assessment of racehorse behaviour is needed.

Wearable Inertial Measurement Units (IMUs) use orientation and acceleration to interpret activity (review by Mathie *et al.* [121, 133-135]) and have been used in healthcare to detect Parkinson's disease [136] and in biology to study movement of marine, terrestrial and airborne animals [125, 126, 137, 138]. Movement is recorded as acceleration signals and analysed by machine learning tools to identify patterns that can be used to discriminate behaviours. Algorithms have been developed and validated to classify behaviour in dogs with excellent agreement (>0.90) [131], detect lameness in sheep with an accuracy of 82% [122], monitor lying behaviour in cattle [139] and predict foal birth in prenatal broodmares [124].

Effective remote monitoring of horses could be used in racing stables to detect subtle changes in behaviour. The objective of this study was to determine whether intensive methods of assessing horse behaviour, such as constant observation, can be replaced by IMUs. We aimed to develop an algorithm to classify horse behaviour and (1) determine the accuracy and precision of the behavioural events predicted by IMUs; and (2) validate their use to quantify posture and behaviour, specifically, in detecting static orientations of standing and recumbency as well as dynamic movements of walking, pawing and weight shifting.

2.3 Materials and methods

2.3.1 Study Population

Twenty-five thoroughbred horses were recruited; from the University of Melbourne Equine Centre (U-Vet) (n=6) and from one Victorian registered racehorse trainer (n=19). The study was approved by the University of Melbourne Animal Ethics committee (reference number 1814523). Six horses recruited from U-Vet were selected based on their likeliness to display the behaviours of interest (mean age 3.7 years, sd 2.2 years). The developed algorithm was tested on video recordings of nineteen horses that met the inclusion criteria; (1) Thoroughbred racehorse; (2) ≥ 3 years and (3) deemed fit for race training. The mean age of the racehorses in training was 3.8 years (sd 1.07); females: n=3; geldings: n=13; colts/stallions: n=3.

2.3.2 Sample size analysis

Estimations of sample size were based on previous published validation of motion sensors in dairy cattle [140]. At least 13 horses were required to determine whether a correlation coefficient differed from zero with an expected correlation coefficient between video observations and algorithm predictions of 0.7 or greater, with a power of 80% and a level of significance of 0.05. We used the statistical software G*Power correlation (bivariate normal model) to estimate required sample size [141].

2.3.3 Accelerometer data

Horses were equipped with two 500Hz nine axis IMU (length: 40mm; width: 28mm; depth: 15mm; mass: 12g; accelerometer range: ± 16g; gyroscope range ± 2000 degree per second (°/s); Vicon, Auckland, New Zealand). Sensors were positioned on the lateral side of both metacarpi, placed in a horse boot with a customised Velcro pouch (Supplementary Figure S1). We chose to place the IMU device on forelimbs to detect pawing and weight shifting behaviour.

The duration of data collection was limited to the battery life of the sensors (<5 hours) and therefore collected at three separate time points: in the morning (7:00am - 11:00am), afternoon (2:00pm – 6:00pm) and evening (6:00pm – 11:00pm) to provide opportunity for horses to display a wide range of behaviours (for example: walking, standing, pawing, weight shifting, lying down). These time frames were dictated by stable activities and horse availability. A mobile application (IMeasureU, Vicon) was used to start and stop data capture.

2.3.4 Video analysis

Video footage was recorded using a video camera (TECHview 1080p) mounted to a bracket that was securely fastened to the stable wall (stable size 360mm x 370mm) in one corner at a height of 2.40m, which allowed for a view of the entire stable (101°). The mean footage for each horse was 224 minutes (sd 37.6). The sensors were tapped by hand multiple times in view of the camera to provide a time stamp that was used to synchronize the video footage with the IMU data [121].

2.3.5 Data processing and algorithm development

Data processing and algorithm development was conducted using MATLAB version R2018a (Mathworks, Natick, MA).

Sensor data was extracted from the IMUs and imported into MATLAB. A custom designed MATLAB script was used to process the data. The z and x axis from the sensor of the left forelimb were multiplied by negative one to convert data equivalent to that of the right forelimb.

Data was collected from hospital-admitted horses (n=6) and used to train the algorithm (training data). The video recording was used to manually label time frames from sensor data to a behaviour class. The time stamp allowed one observer (KA) to coordinate IMU time with video footage. For each movement on IMU, the start time was recorded and labelled appropriately to the behaviour observed on video in one second intervals [134].

This time interval was chosen because it has previously been reported to be sensitive to detection of movement for most behaviour classes [125].

We defined six basic non-overlapping behavioural classes: lying down left, step, pawing, stand, weight shift and lying down right (Table 2:1). Each behavior was encoded: 1= laying down left; 2= step; 3= pawing ground; 4 = standing; 5= weightshifting; 6= laying down right. Each classified behavioural observation was allocated to a row number which was found by multiplying the time in seconds by 500 (because the data was recorded at 500hz; each second contained 500 points). Five features were calculated using a moving window calculation for the time interval across the entire array; including mean, median, minimum, maximum, standard deviation. The five features were calculated for each of the eight columns, consisting of six acceleration signals (acceleration x,y,y; gyroscope x,y,z) and an additional two resultant vectors calculated using the Pythagorean method. These calculations were associated to a labelled behaviour and used to train machine learning classifiers. Each classifier was evaluated using 10-fold Cross-Validation and the accuracy of each was determined [127].

A confusion matrix describing the performance of the algorithm at classifying behaviours was generated for the training dataset [127]. The positive predictive value was the proportion of predictions that were correct, and the true positive rate was the percentage of actual movements that were predicted correctly

Table 2:1 Description of broadly classified behaviours observed on video in n=25 Thoroughbred racehorses.

Behaviour	**Value**	**Description**
Laying down left	1	The horse is in sternal or lateral recumbency with its left side down for >3s, including behaviour such as rolling.
Step	2	Movement of hoof from one location to another one
Pawing ground	3	Continuous action of digging with one limb with the horse otherwise stationary.
Standing	4	Standing on both forelimbs without unloading or moving (≥1s)
Weightshifting	5	Action of unloading weight from one limb to another (not necessarily taking a limb off the ground), often continuous between left forelimb and right forelimb.

Laying down right	6	The horse is in sternal or lateral recumbency with its right side down for >3s, including behaviour such as rolling.

2.3.6 Validation

The developed algorithm was used to predict behaviour in n=19 horses in race training (unlabeled test dataset) using patterns recognised from the previously labelled training dataset [126, 127]. The predicted behaviours from the algorithm were blinded to the observer (KA), who manually classified behaviour per second from the video to one of the six defined behaviours. If behaviour could not be recorded because the limb of interest was obscured, the data was recorded as missing. Abnormal behaviour, i.e. if activity could not be classified (for example: spook to the right), was also recorded.

An internal clock drift in the IMU was corrected using a time stamp to detect the difference between true and IMU time. Each sensor was manually calibrated at consistent intervals based on the correction factor determined. Since the time resolution of the device was 1 s, the correction was implemented in the data as occasional repeats of the same time, or occasional 1-s skips as previously reported [142].

The transitional period between stand and step was frequently classified as weight shifts (LF 55.8% (4087/7316) and RF 50.7% (3623/7143) of total weight shifts). To account for this misclassification a variable to exclude weight shift movements predicted before and after step behaviour was generated. A period of 5 seconds following the transition of lying down to standing was not used in validation analysis because behaviour displayed could not be defined.

2.3.7 Statistical Analysis

Statistical analyses were conducted in Stata IC version 15 (2017, StataCorp LLC, College Station, Texas USA). The unbalanced ratio of behaviours meant the accuracy for all classifications was excellent and not informative to the overall performance of the algorithm. Therefore, we evaluated performance of the algorithm based on its ability to correctly predict individual behaviours, including sensitivity and precision, as previously done by Martiskainen *et al.* [130]. Positive predictive value (PPV), sensitivity (true positive; †TP), and precision were determined using a Bland–Altman analysis to assess the accuracy of the algorithm to predict horse behaviour in the test data (n=19), by comparing the video observation to the behaviour predicted. Sequence index plots were generated using Stata. Sequence analysis was conducted in R using the package TraMineR [143]. The duration of each behaviour was calculated as a mean of observed behaviour from each behaviour. For lying down behaviour, the duration percentage was a mean calculated from only those horses that did lie down.

2.4 Results:

2.4.1 Training data

Table 2:2 presents a confusion matrix describing the performance of the behaviour classifier algorithm, with an accuracy ranging for each behaviour from 90-100%, and an overall accuracy of 96% on the training data. Lying down left (1) had the highest PPV, followed by Lying down right (6), Standing (4), Pawing ground (3), Weightshifting (5) and Step (2), whereas stepping (2) and pawing (3) had the lowest sensitivity.

Table 2:2 Performance of the trained classifier (algorithm) on the training dataset of hospital-admitted Thoroughbred horses (n=6)

		Observed behaviour						
		1	2	3	4	5	6	‡PPV (%)
Predicted behaviour	1	158						100
	2	1	179		5	12	1	90
	3	2		113		3		96
	4		28	2	1539	10		97
	5	2	15	25	3	559	2	92
	6	1	1			12	85	99
	†TP(%)	96	81	80	>99	96	97	

1= Lying down left; 2= Step; 3= Pawing; 4= Standing; 5= Weight shifting; 6= Lying down right; †TP= true positive; ‡PPV= Positive Predictive Value.

2.4.2 Test data

Excluding data that could not be observed, other behaviours (not classified within the six classifications) and behaviour detected in the second before and second after 'step' from analysis, video observation detected standing as the most dominant behaviour pattern in the test dataset (n=19 horses; mean percent: LF 89.6, sd 7.6 and RF 89.3, sd 7.5) followed by laying down (mean percent: LF 4.5, sd 8 RF 4.7, sd 8.2 and LF 1.9, sd 3.9 RF 1.8, sd 3.9; LDL and LDR, respectively) and step (mean percent: 3.0, sd 1.8 for both LF and RF). Over the complete observation period, 57% of horses (11/19) lay down for a period of more than three seconds, all of which occurred during the afternoon (between 3pm and 5pm) or evening (after 8pm) (Figure 2:1). The mean lying time for all horses was 14.5 minutes (sd 19.4; range 0-70.9) One horse lay down for 27.5% of the observational period (70.9/257.6 minutes).

Table 3 describes the summary data recorded from 19 individual horses. Data for each horse from the test dataset is available in Supplementary Table S3.

It was likely (>60%), for all behaviours except weight shifting, that the behaviour state observed was followed by the same behaviour state in the next recorded second (Supplementary Table 2:7). Weight shifting was most likely followed by standing (0.47) or further weight shifting (0.42). Calculations of transition rates when repeated events in sequence were ignored showed a high probability that a step was followed by standing (0.90) and weight shifts were followed by standing (0.81). The longitudinal characteristics of sequences between horses is illustrated by Figure 2:1, which can be used to interpret how individual horses change behaviour state over time.

Table 2:3 Summary information about the behaviour sequences of both left and right forelimb of n=19 individual Thoroughbred horses in race training monitored in a stable.

	Left limb observed		Right limb observed	
	Mean (sd)	Min, max	Mean (sd)	Min, max
[1]Total length of observation (seconds)	12093 (2149)	8512, 16214	11886 (2095)	8415, 14763
No. of transitions	448 (165)	231, 982	431 (140)	230, 725
States observed (max=6)	4 (1)	3, 6	4(1)	3,6
[*]Duration of each behaviour (%)				
Laying down left (8/19)	4.4 (8.2)	0, 29.8	4.7 (8.5)	0,29.9
Step	3.5 (2)	1.5, 9.8	3.5 (2)	1.8, 9.4
Pawing ground	0.1 (0.1)	0, 0.2	0.2 (0.2)	0, 0.7
Standing	89 (7.7)	66.6, 97.9	88.7 (7.6)	66.7, 97.3
Weightshifting	1.2 (0.7)	0.1, 2.6	1.3 (0.8)	0.1, 3.5
Lying down right (4/19)	1.9 (3.9)	0, 11.2	1.8 (4.0)	0, 12.4
[†]Entropy	0.22 (0.1)	0.06, 0.42	0.23 (0.1)	0.08, 0.42

[‡]Complexity	0.09 (0.1)	0.04, 0.16	0.009 (0.02)	0.05, 0.13

[1]The total length of observation excluding time when horse could not be observed, [*]The percent of time spent in each state of behaviour during the observational hours, which include weight shifts before and after step which were removed for Stata analysis;[†] a measure of diversity of states that compose the sequence, based on time spent at different states (where 0 = one behaviour observed and 1 = equal time spent at each state), [‡] combines entropy and number of transitions.

The total agreement between predicted and observed behaviours is shown in Table 2:4. These numbers were used to calculate the precision and sensitivity.

$$\text{Sensitivity} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

Table 2:5 presents the sensitivity (proportion of the positive cases that were predicted positive) and precision (proportion of positive predictions that were correct) of the algorithm to predict each behaviour classification, stratified by limb. Overall, the precision varied from 0.29 (weight shifting) to 1 (laying down), and sensitivity varied from 0.50 (pawing ground) to 1 (laying down). Standing and laying down were predicted with a precision of >0.99 in both left and right limbs. Compared with video observations, the predictions from the algorithm showed an overall misclassification of 2.4%. Excluding standing (the most prevalent behaviour), the overall misclassification was 15.8%. Pawing ground was most often misclassified as step (LF 27% and RF 50% of true positives). Weight shifting was most often misclassified as either standing or step (LF 46%; RF 38% of true positives). Step behaviours were predicted with excellent precision and sensitivity.

Table 2:4 Total agreement between predicted and observed behaviour n=19 Thoroughbred stabled horses

	Observed													
Predicted	Laying Left		Step		Pawing ground		Standing		Weight shifting		Laying Right		Total	
	LF	RF	LF	RF	LF	RF	LF	RF	LF	RF	LF	RF	LF	RF
Lying Left	**11477**	**11466**	13	0	0	0	0	0	2	2	1	3	11481	11471
Step	1	13	**5805**	**4115**	31	51	1113	735	477	333	1	0	7440	5247
Pawing ground	0	1	29	8	**79**	**56**	13	3	4	5	1	0	128	73
Standing	0	1	441	411	3	0	**196405**	**181496**	479	461	0	0	197330	182369
Weight shifting	2	9	171	1061	1	2	2548	3604	**1127**	**1644**	1	0	3848	6320
Laying Right	1	0	0	0	0	0	0	0	0	0	**4950**	4278	4950	4278
Total	11494	11490	6447	5595	114	109	200079	185838	2089	2445	4954	4281	**225177**	**209758**

Table 2:5 Results of the Bland-Altman analysis assessing sensitivity and precision and their respective 95% confidence intervals (95%CI) of behaviours between video observation and algorithm prediction of behaviour in n=19 stabled Thoroughbreds, stratified by limb.

	Sensitivity (95%CI)			
Behaviour	Left Forelimb	95%CI	Right Forelimb	95%CI
Laying down Left	0.999	0.998 – 0.999	0.998	0.997 – 0.998
Step	0.90	0.893 – 0.907	0.861	0.852 – 0.870
Pawing	0.693	0.602 – 0.771	0.496	0.404 – 0.587
Standing	0.981	0.981 – 0.982	0.982	0.982 – 0.983
Weightshifting	0.539	0.518 – 0.561	0.615	0.595 – 0.635
Laying down Right	0.999	0.998 – 1.00	0.999	0.998 – 0.998

	Precision (95%CI)			
Behaviour	Left Forelimb	95%CI	Right Forelimb	95%CI
Laying down Left	0.999	0.999 – 1.00	0.999	0.999 – 1.00
Step	0.780	0.771 – 0.790	0.791	0.781 – 0.801
Pawing	0.617	0.530 – 0.680	0.736	0.626 – 0.824
Standing	0.995	0.995 – 0.996	0.994	0.994 – 0.995
Weightshifting	0.292	0.279 – 0.307	0.341	0.326 – 0.355
Laying down Right	1.00	.	1.00	.

Figure 2:1 Sequence index plot demonstrating changes in behaviour states over time of thoroughbred racehorses in training (n=19). Changes in behaviour state are shown by changes in colour.

During the 4 hours of continuous monitoring, test horses displayed behaviours that did not fit into one of the six classifications included in the development of the algorithm. Such observed behaviours included: shake, roll, kick, lift leg, itch leg. These behaviours were allocated to “other behaviours” (n=925) and excluded from calculations of agreement.

2.5 Discussion:

Using an inertial measurement device, an algorithm was developed and validated capable of accurately identifying horse behaviour, and therefore replacing current methods of observation and enabling continuous monitoring. Using a supervised learning algorithm, we achieved excellent sensitivity for four behaviours: laying down left, laying down right, step and standing; and fair to good sensitivity for pawing ground and weight shifting. The overall performance of the device suggests it could be implemented, developed into a system able to monitor Thoroughbred racehorses in their stables, and over time may allow detection of subtle changes in behaviour patterns.

The horses in our study spent the greatest time standing (89%) and least time pawing the ground (<1%). A previous study observing stabled horses for 72 hours, (divided into 12 hour segments over 6 days) found the majority of time was spent standing (55%) and ‘grazing’ (24%) [144]. Our findings are similar as we did not individually categorise eating behaviours. All horses were recumbent only after 3pm, similar to previous research that observed horses were only recumbent during the night [144, 145]. Horses have been observed to spend most of their recumbent time in sternal recumbency rather than on their side [144]. We were unable to differentiate sternal and lateral recumbency, however we did determine whether horses were lying on the left or right side.

The sensitivity for detecting laying down and standing in this study was high (>98%), comparable to previous validation studies in cattle and horses (≥98%) [145-147]. Fair to good sensitivity was recorded for predicting pawing ground, with the right forelimb sensor correctly predicting only 50% of observed pawing bouts. Pawing movements can be brief so mismatches between the sensor and human observation time recordings are more likely than for behaviours of a longer duration. Similar validation studies found that when slowing the recordings down to review disagreements the device predicted correctly, and the observer was incorrect [131]. The video footage was regularly replayed at 50% of real time to best identify these swift movements but the recording of behaviour every second allowed little room for error. Recently, a small study using observation and convolutional neural networks (CNN), found the largest contributors to a lower accuracy were the misclassifications of pawing ground, rolling and flank watching [148]. These behaviours were only performed by one horse during the collection of training data and therefore made up only 2.4% of the dataset. The infrequent display of pawing ground limits training opportunities for classifiers. Additionally, individual pawing patterns were diverse, producing distinctive acceleration patterns for each horse unlike the repetitive movement pattern observed during walking or lying down. Therefore, larger input datasets are required for training algorithms to recognise pawing.

We placed the motion sensors on the forelimbs. This position was previously shown to achieve the highest accuracy for discriminating between equine gaits, including walk, trot and canter [149]. Placement of a motion sensor around a cows' neck has been successful for recognising rumination and eating behaviours [130, 150-152], but sudden head movements can disrupt recognition of other behaviour patterns and collars may move freely, independent of animal movement [151]. Additionally, sensors placed on the neck have difficulty differentiating standing and sternal recumbency because the orientation of the axis does not change [130]. Forelimb placement can differentiate these postures using the perpendicular difference between x and y axis [153]. Robert *et al.* placed the motion sensor on the hindlimb of cattle and reported good agreement with video for lying and standing (99.2% and 98%, respectively) [146]. However, walking classification accuracy was significantly ($p<0.01$) lower (67.8%). Sensors placed on the ear, collar and leg yielded different prediction accuracy for lameness in sheep [122]. Leg mounted sensors cannot differentiate standing from grazing, unlike ear mounted sensors which can distinguish non-grazing standing with a 96% prediction accuracy. Additional sensors strategically placed could be utilised to detect a movement pattern unique to one behaviour that discriminates it from others. Although multiple sensors would not be applicable for practical reasons in a racing stable, further investigations would be required to determine if different sensor placement could improve prediction accuracy. Future work could include classifying more behaviours that are performed by racehorses and associated with stress or discomfort like: eating behaviour or time spent in sternal or lateral recumbency.

Weight shifting was difficult to classify, compared to behaviours that occurred over longer periods of time without change. The high probability of a transitions between step, standing and weightshifting (Table 2:7) may account for the difficulty in determining weight shifts as an independent movement or as a transitional behaviour. In attempt to minimise this, we excluded weight shifts one second before and one second after step behaviour, although this time period may not effectively account for all transient detections of weight shifts (such as before and after pawing ground). Therefore, it may be necessary to make improvements to the algorithm to increase sensitivity for weight shifting as an individual movement. One possible solution would be to account for the behavioural states within a time period before and after the prediction. For example, weightshifting only be predicted if the two seconds preceding and following standing.

The algorithm showed good sensitivity to correctly predict observed weight shifts (LF 54% and RF 61%), however precision was poor (30-34% of positive predictions were correct). Misclassified weight shifts predicted by the algorithm were most commonly recorded by the observer as standing. This may be due to the high sensitivity of the IMU system to detect small accelerations (sampling rate: 500Hz) compared with the temporal resolution of the human eye (15-20Hz) [154]. It is difficult to determine true sensitivity when the gold standard is subjective and objective tools have shown higher sensitivity to detect clinical signs prior to human diagnosis [84, 95]. Force plates could be used to objectively quantify weight shifts and validate the sensitivity of the motion sensor to detect and classify subtle weight shifts potentially overlooked by observation.

Our study showed higher sensitivity and precision than studies using 10 second intervals, except for walking which was comparable, most likely because walking can be a continuous behaviour and easily averaged over 10 seconds. We chose to classify behaviour at a time-sampling interval of one second attempting to obtain the highest sensitivity for subtle movements. Intervals of 1 second were used in dogs to achieve >95% accuracy for 6 of 8 behaviours [131]. The combination of sampling rate and time interval used can influence the accuracy of detecting horse behaviour [148]. Eerdekens *et al.* showed an increase of time interval from 0.6s to 1.2s resulted in significant improvements to predictive performance ($p \leq 0.05$) [148]. Robert *et al.* [146] compared the accuracy of 3, 5 and 10 second intervals to determine behaviour in cattle and found agreement between sensor and observation was higher in 3 (98.1%) and 5 second intervals (97.7%) compared to the 10 second intervals (85.4%). High variation in readings recorded over 10 second intervals were thought to reduce accuracy.

This study had some limitations. Only one observer manually labelled the defined behaviours from video observation. Future studies should consider additional observers when labelling data to reduce errors associated with subjectivity and human perception. However, previous observational studies have assessed both intra and inter reliability with positive results. Ledgerwood *et al.* reported intra reliability to be good for one observer to record behaviour in cattle and DuBois *et al.* (2015) found there was no significant difference between observations of lying down in horses recorded by two observers [145, 147]. Secondly, the accelerometer and gyroscope used in this study had a battery life of less than 5 hours. Although adequate for validation purposes, this would be a limiting factor for the clinical applicability of its purpose as a 24hr monitoring system. Previous studies showed accuracy was only reduced 5% when sampling rate was reduced from 200Hz to 25Hz [148]. A reduced sampling rate would improve battery performance, but this would require further algorithm development. Lastly, we used six horses in the development of the algorithm, thus increasing the number of horses and types of behaviours may improve predictions.

IMUs have the potential to objectively quantify horse behaviour and detect changes over time. Although the association between each behaviour and musculoskeletal injury is still unknown, the ability to document behaviour over time may allow the identification of patterns indicative of pathology. The overall performance of the algorithm we developed to determine horse activity was good. However, further algorithm development is required such as distinguishing distinct bouts of weight shifting from those that are a transition to another behaviour.

2.6 Supplementary:

Figure 2:2 Placement of sensor on forelimb with attention to direction of axes

Table 2:6 Summary information about the left forelimb behaviour sequences of n=19 individual Thoroughbred horses being monitored in a stable.

Horse ID	Sequence length (seconds)	No. of transitions	*Duration of each behaviour (%)						†Entropy	‡Complexity
			LL	Step	Pg	Stand	Ws	LR		
1	10115	231	0.0	1.5	0.0	97.9	0.5	0.0	0.06	0.04
2	14410	982	0.3	9.8	0.2	87.4	2.4	0.0	0.26	0.13
3	14394	554	9.2	2.0	0.2	86.1	2.6	0.0	0.30	0.11
4	10844	444	0.0	6.0	0.2	87.6	1.2	5.0	0.28	0.11
5	11507	283	0.0	3.7	0.2	96.0	0.2	0.0	0.10	0.05
6	13194	469	1.8	2.3	0.0	86.4	1.4	8.2	0.31	0.10
7	10618	607	0.0	4.7	0.0	92.7	2.6	0.0	0.17	0.10
8	13698	288	7.6	2.2	0.2	88.9	1.1	0.0	0.25	0.07
9	17205	478	0.0	2.2	0.0	85.7	0.9	11.2	0.28	0.09
10	14449	404	4.3	2.4	0.0	92.1	1.2	0.0	0.20	0.07
11	14690	483	20.0	3.0	0.0	76.4	0.6	0.0	0.37	0.11
12	9426	522	0.0	6.5	0.0	92.5	1.0	0.0	0.16	0.10
13	15459	374	0.0	3.2	0.0	85.0	0.6	11.2	0.29	0.08
14	14190	531	0.0	3.3	0.0	95.5	1.2	0.0	0.12	0.07
15	14423	312	10.5	2.3	0.0	86.6	0.6	0.0	0.27	0.08
16	15450	276	29.8	3.4	0.0	66.6	0.1	0.0	0.42	0.09
17	10763	386	0.0	3.4	0.0	95.2	1.4	0.0	0.12	0.07
18	14899	333	0.0	2.3	0.0	96.6	1.1	0.0	0.09	0.05
19	14405	551	0.0	2.6	0.0	95.7	1.7	0.0	0.11	0.07
mean	13376	448	4.4	3.5	0.1	89.0	1.2	1.9	0.22	0.08
sd	2162	165	8.2	2.0	0.1	7.7	0.7	3.9	0.10	0.02
min	9426	231	0.0	1.5	0.0	66.6	0.1	0.0	0.06	0.04
max	17205	982	29.8	9.8	0.2	97.9	2.6	11.2	0.42	0.13

*The percent of time spent in each state of behaviour during the observational hours, which include weight shifts before and after step which were removed for Stata analysis; † a measure of diversity of states that compose the sequence, based on time spent at different states (where 0 = one behaviour observed and 1 = equal time spent at each state), ‡ combines entropy and number of transitions.

Table 2:7 Transition rates, or the probability of transition from one behaviour state to another during the observational hours of n=19 stabled Thoroughbred horses.

	Laying left	Step	Pawing	Standing	Ws	Laying right
Laying left	**1.00**	0.00	0.00	0.00	0.00	0.00
Step	0.00	**0.61**	0.00	0.35	0.04	0.00
Pg	0.00	0.18	**0.62**	0.13	0.07	0.00
Standing	0.00	0.01	0.00	**0.98**	0.01	0.00

Ws	0.00	0.10	0.00	0.47	**0.42**	0.00
Laying right	0.00	0.00	0.00	0.00	0.00	**1.00**

Transition rates of left forelimb calculated from total observations.

	Laying left	Step	Pawing	Standing	Ws	Laying right
Laying left	**0.00**	0.13	0.00	0.33	0.53	0.00
Step	0.00	**0.00**	0.00	0.90	0.10	0.00
Pg	0.02	0.47	**0.00**	0.34	0.17	0.00
Standing	0.00	0.68	0.00	**0.00**	0.31	0.00
Ws	0.00	0.18	0.00	0.81	**0.00**	0.00
Laying right	0.00	0.00	0.00	0.00	1.00	**0.00**

Transition rates of left forelimb when states go direct to next event (repeated events in sequence are ignored).

	Laying left	Step	Pawing	Standing	Ws	Laying right
Laying left	1.00	0.00	0.00	0.00	0.00	0.00
Step	0.00	0.62	0.00	0.34	0.04	0.00
Pg	0.00	0.19	0.64	0.09	0.08	0.00
Standing	0.00	0.01	0.00	0.98	0.01	0.00
Ws	0.00	0.10	0.00	0.43	0.47	0.00
Laying right	0.00	0.00	0.00	0.00	0.00	1.00

Transition rates of right forelimb calculated from total observations.

	Laying left	Step	Pawing	Standing	Ws	Laying right
Laying left	0.00	0.15	0.00	0.38	0.46	0.00
Step	0.00	0.02*	0.01	0.88	0.09	0.00
Pg	0.00	0.54	0.00	0.24	0.22	0.00
Standing	0.00	0.66	0.00	0.02*	0.32	0.00
Ws	0.01	0.19	0.00	0.80	0.00	0.00
Laying right	0.00	0.00	0.00	0.00	1.00	0.00

Transition rates of right forelimb when states go direct to next event (repeated events in sequence are ignored). *repeated in sequence when HorseID changes.

Chapter 3: Investigating whether facial grimace scores can predict lameness in Thoroughbreds during trot up examinations

3.1 Introduction:

Whilst overt signs of lameness in horses can be readily identified and therefore promptly addressed, subtle signs are more challenging to detect [99]. Many orthopaedic injuries are the result of accumulated microdamage over time thus there is a spectrum of injury ranging from subclinical (i.e., no appreciable clinical signs) to severe lameness. In order to minimise the likelihood of progression to overt fracture, more sensitive methods for evaluating the accumulation of fatigue injury are required. When microdamage accumulates faster than it can repair, the integrity of the bone will eventually be overcome resulting in gross injury such as fracture [11]. Sudden onset of moderate to severe lameness that rapidly resolves is a key indicator of stress fracture [155]. But, since horses that fracture do not typically demonstrate clinical signs of an impending problem, the accurate detection of horses at risk is challenging.

Equine lameness is defined as an alteration of the normal gait due to a functional or structural disorder of the loco-motor system, and is commonly attributed to orthopaedic pain [71, 85]. Lameness is the most commonly reported reason for failure to train and reduced racing performance, affecting 52% of the racehorse population [156, 157]. Numerous studies have revealed the high incidence of horses with gait asymmetries not detected by owners or trainers [79, 99, 100]. Lameness is most commonly assessed using subjective visual evaluation of (a)symmetry between left and right side during the gait at trot [73]. Despite the use of standardised scales, reproducibility between multiple scales including AAEP, numerical rating and verbal rating, is low [76, 84]. These findings highlight some of the challenges in the detection and assessment of horses for lameness. Objective tools can be used to evaluate gait symmetry and improve reliability [84]. Kinetic gait analysis has shown potential to detect reduced force prior to subjectively observed lameness [95]. Wireless inertial sensor units for evaluation of lameness on all limbs over ground have proved repeatable for use in a clinical setting [80]. However, for everyday use and in stabling environments better routine methods of detection are required. Following the detection of gait asymmetry, the interpretation of whether it represents orthopedic pain requires further work up and can be confirmed using diagnostic analgesia [158].

Scales that grade a horses facial expression have been developed to objectively evaluate pain from various sources [12, 67, 117]. The Horse Grimace Scale (HGS) was validated to quantify pain in horses post-castration surgery [12, 113] and also applied to assess horses after orthopaedic surgery [159]. Facial expressions in ridden horses (FEReq) differentiate lame and sound horses [16, 117]. Horses that were lame displayed certain behavioural markers more often than sound horses, and four markers were exclusively observed in lame horses [118]. However, behavioural and facial parameters are yet to be assessed in non-ridden horses, for example during an in hand trot up lameness examination.

The objective of this study was to assess the validity and clinical applicability of the HGS and FEReq for differentiating between lame and 'fit to race' Thoroughbred horses during a standard trot up examination. We hypothesised that high facial pain scores would be associated with lameness above that considered appropriate for racing.

3.2 Materials and Methods:

A total of 39 Thoroughbred horses were recruited: 13 horses from the University of Melbourne Equine Veterinary Clinic admitted to the clinic for lameness and poor performance, and 26 horses, deemed fit for racing, from the training yard of one Victorian registered trainer. Information for each horse including age, sex and racing history, were obtained from the official repository for racing information in Australia (Racing Australia; https://www.racingaustralia.horse/). The study was approved by the University of Melbourne Animal Ethics committee (reference number 1814577.1).

3.2.1 Lameness examination

Objective lameness assessments were performed using a body-mounted inertial sensor system (Lameness Locator®. Equinosis, St Louis, Missouri, USA). Each horse was instrumented with two single-axis acceleration sensors and one single-axis gyroscope sensor (3.8 X 2.5 X 1.3 cm; 30g). Sensors were turned on with a magnetic switch prior to attachment. One acceleration sensor was placed on the poll using a head bumper attached to the halter. The second accelerometer was placed on the tuber sacrale attached with 3M dual lock tape and secured with additional tape over the top. The gyroscope sensor was placed on the dorsal surface of the right forelimb pastern. Total instrumentation time was < 2 minutes. Data obtained by the three sensors were wirelessly transmitted in real time (8-bits at 200 Hz) to a tablet, and a series of motion analysis algorithms previously developed were automatically performed to evaluate gait and provide diagnosis [90, 160, 161]. Lameness reports were developed using asymmetry recorded by the LL for the following variables: maximum head displacement (difference between maximum heights of the head after stance of forelimbs); minimum head displacement (difference between minimum heights of head during stance of forelimbs); total difference head; maximum pelvis displacement (difference between maximum heights of the pelvis after stance of hindlimbs); minimum pelvis displacement (difference between minimum heights of pelvis during stance of hindlimb) [80]. Movement symmetry was determined by calculating differences in minimum and maximum values for both head and pelvic displacement during a full stride [80, 88, 90]. The side of asymmetry was indicated by the positive or negative sign, positive assigned to right side. Objective measurements previously prescribed a horse lame when displacement values are greater than thresholds >6mm for head and >3mm for pelvis [162]. Because these thresholds were considered not useful to determine a racehorses fit to race, we used thresholds published by Pfau et al. (>14.5mm for head and >7.5mm for pelvis) [78].

All horses underwent a lameness examination which included an in-hand trot up of over 20 meters (>25 strides) on a hard surface in a straight line. Horses were video recorded from a lateral view (at the centre of the twenty meters) for the duration of the trot up using one high definition camera (Panasonic HC-V770M). All horses wore a bit and bridle during the trot up. Lameness was evaluated subjectively by four veterinarians at the time of data collection based on the scale used by the American Association Equine Practitioners (AAEP), which uses 5 grades to describe severity; where zero is no lameness observed under any circumstances and five is non-weightbearing lame [16, 101, 163]. Trot ups were recorded at the time of lameness examination for hospital-admitted equine patients, and at the time of routine weekly trot up for the horses in training. To randomize selection, every third horse in the training stable was recruited.

3.2.2 Photograph processing

Using the video recording of the gait examination, screenshots were taken at midstance for every footfall between a marked area (5-15m) using the event logging software, BORIS [164]. Screenshots of horses during trot up were captured at mid stance because this is the point of maximum force on the horses limbs [75]. The default random number generator in Stata 15 was used to allocate a number identity to each horse and photograph. Each photograph was documented as stance of left forelimb or right forelimb. Photographs were cropped so that only the head was visible to prevent observers from being biased by the body and the background location [12, 16, 108, 117]. Photographs of poor quality (pixilated, dark), where the entire head was not visible, or where obvious tension on the lead was present, were excluded from the collection. For each horse, ten photographs were randomly selected (n=10) and used to create a slideshow (Microsoft PowerPoint), with one image per slide as previously described [165]. Slides were shuffled so there was no order to images to prevent observers deciding on whether a horse was lame which could influence scoring on consecutive photographs.

3.2.3 Pain assessment

Two separate subjective pain scoring systems were employed: (1) the Horse Grimace Scale (HGS) developed by Dalla Costa *et al* [12], and (2) an ethogram using facial expression in ridden horses (FEReq) developed by Mullard et al. [117] and adapted by Dyson et al. [16]. Two independent observers, blinded to horse and lameness conducted evaluations twice, 40-50 days apart. Observers had animal welfare backgrounds with limited horse experience. A training manual was provided on how to score pain behaviour (Supplementary File 1). The instructions included emphasis on the importance of recording the presence of each behaviour individually without any regard to the presence of other pain-related parameters which may influence one to observe pain behaviour not present. For observations where observers could not see or grade, rather than speculating, observers were instructed to score as “cannot tell”.

The HGS includes six facial action units (FAUs): stiffly backwards ears; orbital tightening; tension above the eye area; prominent strained chewing muscles; mouth strained and pronounced chin; strained nostrils and flattening of the profile. For each photograph the observer scored each individual FAU using a 4-point scale (0= not present, 1= moderately present, 2= obviously present, or CT= cannot tell) [12]. Additionally, the observer was asked to make an overall pain judgement using a simple descriptive scale (SDS; no pain, mild, moderate, severe) proven to be highly repeatable [12, 83, 116].

We modified the FEReq ethogram to exclude two behaviour parameters (position of head and bit), as we regarded this irrelevant for trot up and not comparable between horses due to influence from tension of handler [16, 117]. Observers were to confirm the presence or absence of behaviours. A pain score, blinded to observers, was allocated to each behaviour, previously determined to correlate to severity of pain (Supplementary File 1) [16].

3.2.4 Statistical Analysis

Cohens Kappa (κ) was used to measure the proportion of agreement between raters [166, 167]. Interobserver reliability was calculated for first observations, second observations and all observations. Values were classified as <0.20 poor agreement, 0.21 to 0.40 fair, 0.41 to 0.60 moderate, 0.61 to 0.80 good, and >0.80

very good [168]. Intra-observer reliability was calculated to determine agreement within each observer between first and second observations on the same images. The HGS parameters were scored into four different categories, therefore to reflect degree of disagreement, such that disagreement by one scale has less effect than disagreement by two scales, the kappa was ordinal weighted [169]. The first observation for observer 1 had 36 horses to score (photographs n=360), compared to all other observations which included 38 horses (photographs n=380).

Logistic regression models for the binary outcome lameness (lame=1; non-lame =0) and linear regression models for the continuous outcome asymmetry values were performed to investigate associations with behavioural predictor parameters, adjusting for clustering on horse to account for multiple observations of each horse. The absolute values of minimum and maximum amplitude for head and pelvis were used as a continuous variable of lameness. Statistical significance was set at p<0.05. Statistical analysis was performed using Stata 15 (StataCorp. 2017. Stata Statistical Software: Release 15. College Station, TX: StataCorp LP).

Data was not nominally distributed therefore median values were used to describe differences in facial scores, Missing values were ignored when calculating the sum of HGS and FEReq scores, which suggests a chance missing data could affect results.

3.3 Results

The horses ranged in age from 2 to 9 years (mean: 3.7 years, sd: 1.5 years) and comprised 13 females, 13 gelded males, and 12 entire males. Horses were allocated as “fit to race” (n=22) or lame (n=16; Table 1) using the inertial measurement units[a] and thresholds by Pfau et al. previously described. Using original thresholds, 8% of horses (3/38) were considered sound. The AAEP graded by veterinarians reported a mean lameness of 1.3 (sd 0.67; range 0-2), including 17 (44.7%) horses graded 2. There were no horses graded >2. Lameness did not differ by sex, with 38.5% (5/13) of lame horses being female, 53.8% (7/13) of lame horses being geldings and remaining 30.7% (4/13) of lame horses being entire males. Lameness did not differ by age, number of race starts or prizemoney.

Table 3:1 Number and mean (standard deviation) of ‘fit to race’ and lame horses from total sample of Thoroughbred horses according to Lameness Locator measurements in trot up examination, stratified by sex.

	Fit to race	**Lame**	**Total**	**p value**
	Mean (sd)	Mean (sd)		
Age	3.4 (1.0)	4.1 (1.8)	3.7 (1.5)	0.282

Race starts	4.3 (6.2)	2.3 (3.3)		0.209
Prizemoney	$111,552 (135,827)	$170,374 (414,096)		0.233
Sex				0.826
Female	8	5	13	
Gelding	6	7	13	
Entire	8	4	12	
AAEP	21	17	38	
AAEP score	0 (0)	1.3 (0.67)		
Lameness Locator	22	16	38	

Table 3:2 Descriptive summary of the inter and intra reliability of two independent observers scoring behaviour parameters twice from photographs of Thoroughbred horses (n=38). Total observations (n=1500).

Behaviour Parameter	Agreement %+	Observations not scored*		Inter-observer Reliability	Intra-observer Reliability
		Observer 1	Observer 2		
Stiffly ears	61.8	0	19	0.57 (0.42,0.68)	0.83 (0.79,0.87)
Orbital Tightening	83.6	0	101	0.76 (0.73,0.80)	0.80 (0.76,0.83)
Tension Above the Eye	54.6	8	111	0.48 (0.44,0.55)	0.60 (0.53,0.64)
Prominent chewing muscles	71.6	2	420	0.45 (0.47,0.59)	0.55(0.22,0.44)
Strained Nostrils	50.9	5	138	0.15 (0.03,0.28)	0.63 (0.34, 0.86)
Mouth Strained	62.7	18	166	0.60 (0.56,0.66)	0.63 (0.54, 0.66)
Overall Pain	46.1	0	19	0.25 (0.18,0.33)	0.45 (0.34, 0.56)
Ears (position)	52.1	0	23	0.69 (0.64,0.73)	0.83 (0.80,0.85)
Stare (normal or intense)	64.4	8	141	0.28 (0.19,0.37)	0.47 (0.29,0.59)
Eye Tension	83.6	0	109	0.76(0.73,0.80)	0.80 (0.75, 0.82)
Sclera (exposed or not exposed)	97.4	1	14	0.72 (0.68,0.76)	0.78 (0.74, 0.81)
Eye shape (Almond or Round)	57.4	0	120	0.23 (0.08,0.37)	0.68 (0.48,0.84)
Tension caudal to eye	76.8	7	93	0.54 (0.49,0.60)	0.66 (0.60,0.69)
Tension above eye	70.1	10	120	0.39 (0.32,0.50)	0.70 (0.92,0.75)
Nostril Tension	56.8	5	150	0.16 (0.06,0.28)	0.56 (0.26,0.79)
Wrinkle Between nostrils	67	163	409	0.14 (0.12,0.35)	0.73 (0.37,0.80)
Wrinkles Ventral to Nostrils	84.4	9	91	0.34 (0.27,0.43)	0.59 (0.50,0.64)
Upper lip	64.7	7	132	0.38 (0.33,0.46)	0.49 (0.40,0.53)
Lower lip	77.1	33	205	0.54 (0.47,0.60)	0.62 (0.50,0.64)
Nose Twisted	95.0	3	17	0.15 (0.07,0.24)	0.64 (0.59,0.68)
Mouth	79.6	2	136	0.67 (0.60,0.74)	0.84 (0.80,0.86)
Tongue	92.2	0	71	0.71 (0.71,0.79)	0.85 (0.82,0.87)
total		281	2805		

+ Agreement when both observers had scored the behaviour, *when observer could not see or score behaviour

Intraobserver reliability was higher than inter-observer reliability for all facial regions. Intra-reliability was very good for tongue (0.85), mouth (0.84), stiffly ears backwards (0.83), ear position (0.83) and orbital tightening (0.80). For all remaining behaviours, agreement within observers was moderate to good.

Mean observer agreement between observers for observations was moderate (0.45; 95% CI 0.36, 0.55). Using the HGS, interobserver agreement was good for orbital tightening, but poor for strained nostrils and overall pain score (<0.40). Inter observer reliability using the FEReq was good for eye tension and poor to fair for eye shape, stare, tension above eye, nostrils tension, wrinkle between nostrils, wrinkles ventral to nostrils, upper lip and nose twisted. All behaviours with good to very good intra observer agreement, had moderate to good inter observer agreement.

Observers were unable to score wrinkles between nostrils in 38% of images, followed by prominent chewing muscles (28%), lower lip (16%), mouth strain (12%) and nostril tension (10%). Observer 2 scored “cannot tell” for 2805 observations, compared to observer 2 who only scored 281 (Table 3:2).

Table 3:3 Association between lame and ‘fit to race’ horses and the HGS and FEReq predictors.

PREDICTOR	N ‘FIT TO RACE’	N LAME	ODDS RATIO	95% CI	P-VALUE	OVERALL P-VALUE
Stiffly backward ears						
Not present	324	200	1.00			0.153
Moderately present	410	362	1.43	0.89,2.29	0.136	
Obviously present	122	63	0.84	0.23,3.00	0.784	
Orbital Tightening						
Not present	596	417	1.00			0.448
Moderately present	159	130	1.16	0.56,2.45	0.680	
Obviously present	44	53	1.72	0.68,4.34	0.250	
Tension above the eye						
Not present	240	243	1.00			0.305
Moderately present	402	280	0.69	0.36,1.33	0.264	
Obviously present	149	67	0.44	0.16,1.27	0.130	
****Prominent Chewing Muscles**						
Not present	344	294	1.00			0.418
Moderately present	243	141	0.68	0.35,1.30	0.242	
Obviously present	39	17	0.51	0.18,1.41	0.195	
***Strained Nostrils**						
Not present	266	239	1.00			0.604

Moderately present	419	300	0.80	0.47,1.35	0.400	
Obviously present	82	51	0.69	0.29,1.65	0.405	
****Mouth Strained**						
Not present	270	552	1.00			0.056
Moderately present	339	180	0.51	0.27,0.97	**0.042**	
Obviously present	156	89	0.55	0.17,1.74	0.307	
***Overall Pain**						
No Pain	405	348	1.00			0.103
Mild Pain	295	177	0.70	0.47,1.04	**0.078**	
Moderate Pain	109	83	0.89	0.46,1.72	0.720	
Severe Pain	38	25	0.77	0.29,2.04	0.594	
Ear Position						
Ears Forward	252	151	1.00			0.350
Erect to side	251	226	1.50	0.74,3.02	0.260	
One ear forward and one ear back	68	46	1.15	0.62,2.13	0.653	
One ear forward and one erect to side	65	63	1.60	0.88,2.90	0.121	
One to side and one to back	84	70	1.39	0.59,3.24	0.453	
Both ears back	117	66	0.94	0.25,3.54	0.923	
***Stare**						
Normal	493	366	1.00			
Stare	285	207	0.98	0.60,1.59	0.930	
Eye Tension						
Not present	580	419	1.00			0.480
Moderately present	152	130	1.18	0.56,2.47	0.661	
Obviously present	39	51	1.71	0.68,4.35	0.253	
Sclera						
Not exposed	813	582	1.00			
Exposed	23	54	3.33	1.03,10.79	**0.045**	
***Eye Shape**						
Round	443	272	1.00			
Almond	345	320	1.51	0.88,2.60	0.137	
Tension caudal to eye						
Not present	295	295	1.00			
Present	484	307	0.62	0.28,1.37	0.236	
***Tension above eye**						
Not present	346	317	1.00			
Present	435	273	0.68	0.35,1.33	0.264	
*****Nostril Tension**						
Relaxed	248	229	1.00			0.476
Open Wide	410	294	0.78	0.45,1.34	0.364	

Tense	103	61	0.64	0.29,1.43	0.277	
*****Wrinkles between nostrils**						
Not present	401	316	1.00			
Present	135	76	0.71	0.36,1.41	0.330	
***Wrinkles Ventral to Nostrils**						
Not present	653	534	1.00			
Present	135	63	0.56	0.27,1.19	0.133	
***Upper Lip**						
Relaxed	424	395	1.00			**0.084**
Tense	251	137	0.59	0.30,1.16	0.126	
Tense and extended	112	42	0.40	0.18,0.90	**0.027**	
****Lower Lip**						
Relaxed	421	395	1.00			
Tense	242	137	0.57	0.26,1.23	0.151	
***Nose Twisted**						
Not present	819	614	1.00			
Present	28	19	0.91	030,2.75	0.861	
Mouth						
closed	555	436	1.00			**0.012**
lips slightly separated: no teeth	111	72	0.83	0.34,2.03	0.680	
lips separated, teeth: no gum	35	15	0.56	0.11,2.93	0.490	
lips separated, teeth and gum exposed	8	4	0.65	0.08,5.32	0.688	
teeth slightly separated: tongue not visible	9	12	1.41	0.35,8.49	0.497	
teeth slightly separated: tongue visible	40	16	0.31	0.15,1.67	0.265	
mouth widely separated but cannot see tongue	5	4	1.15	0.12,9.10	0.971	
teeth widely separated: tongue visible	15	9	0.73	0.11,4.87	0.747	
Tongue						
Tongue in, not seen	720	553	1.00			0.284
Tongue seen inside oral cavity	71	32	0.57	0.21,1.55	0.274	
Tip of tongue protruding, no teeth	13	11	1.12	0.12,10.80	0.919	
Tip of tongue protruding, can see teeth	4	6	1.99	0.18,22.52	0.577	
Large part of tongue out, but teeth not exposed	0	0				
Large part of tongue out and teeth exposed	0	1				

*inter observer reliability of <0.40; ** parameters with >10% missing; ***parameters with >10% missing and inter-reliability of <0.40

Associations between behavioural predictors and lameness are presented in Table 3:3. One association was identified using the HGS; horses with moderately present mouth strain were less likely to be lame (OR 0.51; 95% CI 0.27, 0.97; p=0.042). There were two associations using the FEReq; horses with exposed sclera were more likely to be lame (OR 3.33; 95% CI 1.03,10.79; p=0.045), and horses with tense and extended upper lip less likely to be lame (OR 0.40; 95% Cl 0.18,0.90; P=0.027). Whilst there were no associations between ear position and lameness, three horses with minimal asymmetries (within 6mm for head and 3mm for pelvis) had ear scores of ≤1 in 70% of photographs.

Associations between individual behavioural predictors and lameness severity on a continuous scale are presented in Supplementary Table 3:7. Higher orbital tightening (HGS) and (FEReq) scores were associated with less severity of HDmax lameness, albeit orbital tightening (FEReq) was not significant (p=0.044 and p=0.054, respectively). Higher tension above the eye, for both the HGS and FEReq scores, were associated with an increasing HDmax (p=0.048 and p=0.034, respectively). Presence of wrinkles ventral to nostrils were associated with decreasing HDmin values (p=0.034). As a group of predictors, presence of stiffly ears backwards were associated with an increased severity of HDmin values (p=0.02). Tongue predictors were associated with a decreasing forelimb lameness (HDmin and HDmax) and increasing hindlimb lameness (PDmin and PDmax) (p<0.001), although majority of observations were scored as tongue in, not seen (90.2% observations).

There was no significant difference in the total facial pain between lame (n=64) and 'fit to race' images (n=860) (median, lame 5 and 'fit to race' 6; mean, lame 5.6 ± 3.5 and 'fit to race' 6.1 ± 3.7 standard error) (Supplementary Table 3:5). There was no association between the total HGS score and lameness (median, lame 3 and 'fit to race' 4; mean lame 3.2 ±1.8 and 'fit to race' 3.7 ± 1.9). Horses with HGS total scores of ≥8 were more likely to be in the 'fit to race' group (77%).

3.4 Discussion:

Out of 22 facial expression variables evaluated, only three were associated with lameness at a level that was considered unfit to race [78] and, two of the three associations were in the opposite direction to that hypothesized. When visible, moderate mouth strain and a tense or extended upper lip were observed more commonly in the 'fit to race' group whereas exposed sclera was more often observed in the lame group. In addition, the total facial scores in HGS and FEReq systems were not different between groups. Although consistent within each observer, 45% of the predictors lacked repeatability between observers.

In contrast with the findings of the current study, Dyson *et al.* determined total facial scores were significantly higher for lame ridden horses (median 9; mean, 8.7 ± 0.15) compared to sound ridden horses (median 6; mean 6.1 ± 0.32) based on the six facial groupings ears, eyes, nostrils, mouth, lips and muzzle, and tongue [16]. During development of the FEReq, weightings were allocated to regional facial expressions based on a greater prominence in lame horses; the eyes, ears and nostrils were more heavily weighted than those of the lips, muzzle, or tongue in ridden horses. Exposed sclera was considered a strong indicator of lameness by Dyson et al [118] and this finding was supported by the findings of the current study, where it occurred more often in horses categorized as lame.

Whilst horses were considered 'fit to race' based on thresholds describing gait asymmetry developed specifically for that purpose, the group did not have lower overall facial pain scores compared with those considered too lame (asymmetric) to race. In the current study, no horses were scored with a lameness severity greater than two by veterinarians using the American Association of Equine Practitioners scale (AAEP) so the range of lameness severity was relatively narrow. This doesn't mean that the FEReq is unable to discriminate lameness of greater severity. The ability for FEReq to determine the presence of musculoskeletal pain was validated previously in lame ridden horses which likely had a greater range of lameness severity, however lameness scores were not stated [16].

A previous study found no significant differences for lips and muzzle between lame and sound ridden horses [16]. In the current study the significant lip and mouth observations were more prominent in the horses categorized as 'fit to race'. These findings suggest that this facial region is insensitive or nonspecific as a predictor of orthopaedic pain or may occur only at very low levels of discomfort. Horses experiencing discomfort from a range of sources may hold their ears backwards and lowered [12, 108, 117]. In the current study no associations were observed between ear position and group, where horses were classified as fit to race or otherwise based on objective thresholds previously published [78]. However, when our population was examined with the grouping more commonly applied to determine the presence of lameness [162] there were three "sound" horses for whom more than 56% photographs (versus 24%) displayed the ears in positions considered representative of comfort. Therefore, it is possible that the high prevalence of lameness in racehorses is a reason facial scoring systems are less sensitive than in other types of horses.

An alternate explanation for the findings in the current study is that some facial expressions occur as a result of other factors such as environmental or handler interactions during the trot up rather than an expression of discomfort. The facial expression is not specific for orthopedic pain [12] and may be confounded by other sources such as gastric ulceration particularly in Thoroughbred racehorses [170], the investigation of which was outside the scope of this study. Previous authors validated the HGS with values shown to increase following surgical intervention [12] and by demonstrating a reduction in pain score in horses following analgesia with acute laminitis [116]. A baseline level of total HGS (2-3 out of a maximum score of 12) in a presumed healthy population of horse's likely accounts for environmental influences on the presence of the measured facial expressions. The total HGS values for both groups in the current study were only slightly elevated from this baseline value. Given the high number of missing values in our population (40.7% photographs had missing HGS values), it is possible our scores underestimated the horses true HGS. Although we tried to account for this in our calculations, could affect the results. Irrespective, it appears the evaluation of facial expressions may not be suited for the discrimination of horses with lameness at least within the narrow range of severity observed in our population.

Thresholds developed by Pfau *et al.* (>14.4mm for Hdmin, and >7.5mm for Pdmin), were determined to better encompass horses "fit to race" [78, 102]. Recent studies have reported 75% of horses fell outside original thresholds [100, 171], suggesting that lame horses can compete successfully and that trainers and riders often consider horses sound in the presence of lameness [99, 172]. Recent discussions have highlighted the need to better understand whether asymmetry is due to a pathological condition and manifestation of pain [158]. For example, asymmetry has been associated with handedness or pre-existing differences in (hind)limb length [158, 173]. The presence of pain can only be confirmed using diagnostic analgesia [118, 163].

A change in expression, particularly of factors observed in horses with greater lameness such as exposed sclera, or an increase in an individual's total score over time, may be more useful in identifying horses that are developing orthopaedic pain than a single assessment. Alternatively monitoring facial scores remotely, for example utilizing video cameras when horses are stabled may be more sensitive for detection of pain as recent studies have demonstrated horses do not as readily display signs of discomfort when they know they are being observed [119].

Interobserver reliability of facial parameters for this study was lower when compared to previous research [12, 16, 116]. Of HGS parameters, our study had lowest agreement for strained nostrils (0.15), also reported to be moderate (0.58) in previous validation studies and the lowest of all HGS parameters [12]. Mean interobserver reliability was only moderate (0.45). The FEReq was previously validated to recognize musculoskeletal pain by one observer and therefore we are unable to compare the interobserver reliability of our findings [16]. The two observers in our study had limited experience with horses and equine behaviour. It is possible this could contribute to different interpretation of facial markers, although studies showed there was no difference in the scoring of observers regardless of professional background [117]. Furthermore, a facial scoring tool should be easily used by individuals, irrespective of experience or background. Intraobserver reliability was higher than interobserver reliability for every facial expression scored, ranging from fair to excellent (0.49 to 0.85). This suggests that although low reliability recorded between observers, the ability for one observer to repeat scores is reliable. Therefore, there is potential for one observer to monitor horses over a period and notice changes in expression.

A large percentage of observations were recorded as “cannot tell” and a substantial difference in observations not scored between observers (2805 versus 281 not scored). This could be explained by one observer foreseeing expressions based on the appearance of other expressions. Observers had most difficulty scoring facial parameters located in the muzzle region, indicated by the number of photographs not scored. Horses in our study were recorded from a lateral position, with images predominantly of the profile view, potentially explaining the high percentage of wrinkles between nostrils not able to be scored. In ridden horses wrinkles between nostrils were also recorded as “cannot tell” in 70% of images [117]. Randomly selected photographs included different angles and some front on images. Prominent chewing muscles scored as “cannot tell” in 28% of images may be explained by these different angles, or the placement of bridle may have obscured this region. The HGS has been validated to score pain in stabled horses but has not been validated on horses during locomotion. In addition to locomotion, these horses were wearing bridles and head piece (with accelerometer attachment) potentially interfering with the observer’s ability to score.

There were a number of limitations to this study. Not all trot ups recorded more than 25 strides (mean 30; range 12-65), as previously recommended by Keegan *et al* [162]. This was due to technical difficulties or horses not cooperating, typical challenges of performing evaluations in a busy training facility. However, previous research has suggested that stride-by-stride variation was not affected by the number of strides collected [80]. It was difficult to maintain a level of consistency between trot ups because trials were on a day to day basis with varying weather, handlers and temperament of horses. A horse that throws the head around repeatedly during a trot up, due to environmental stimuli or handler will result in a wide variation in values for head displacement, although the machine software accounts for removal for some outlier movements. Accelerometers have proved repeatable for measure of asymmetry of pelvic (≥0.93) and head displacement

(≥0.89) [162]. However, a longitudinal study showed racehorses had median weekly differences of 4 to 8mm, indicating potential limitations of using cross sectional data based on IMU gait analysis to classify horses as 'fit to race' [102]. Additionally, cases of symmetric bilateral lameness may be incorrectly evaluated as sound by the Lameness Locator [174].

Variation in pain scores in individual horses may have been affected by the timing of the photo. The nature of taking sampling at one timepoint can easily lead to misinterpretation of behaviour overall. Some horses appear to have eyes closed that was interpreted as orbital tightening, however the photograph could have captured the horse whilst blinking. Video recordings were sensitive for the identification of reduced levels of pain following diagnostic analgesia [163] and may be more clinically applicable to scoring orthopaedic pain in horses during trot up examinations also. Observers in this study scored facial scores based on the presence or absence of parameters irrespective of individual horse personality or behavioural characteristics, and without knowledge of any asymmetrical face or bone structure variations. Knowledge of the individuals normal behaviour may help better define abnormal expression, although this may introduce bias.

In conclusion, few differences in facial pain scores were observed between lame and 'fit to race' horses, and due to the large number of statistical tests performed and the inconsistencies in findings it is possible that those significant findings could be due to chance. Contrary to our hypothesis, facial grimace scales were not suited to the differentiation of Thoroughbred racehorses considered 'fit to race' from those that were lame during in hand trotting and therefore this limits the usefulness of this technique as a potential tool for pre-race examination. However, further longitudinal investigations could be conducted to monitor racehorse lameness and determine if changes in lameness were associated with changes in facial pain scores over time, which may be a more useful judge of impending orthopedic injury.

3.5 Supplementary:

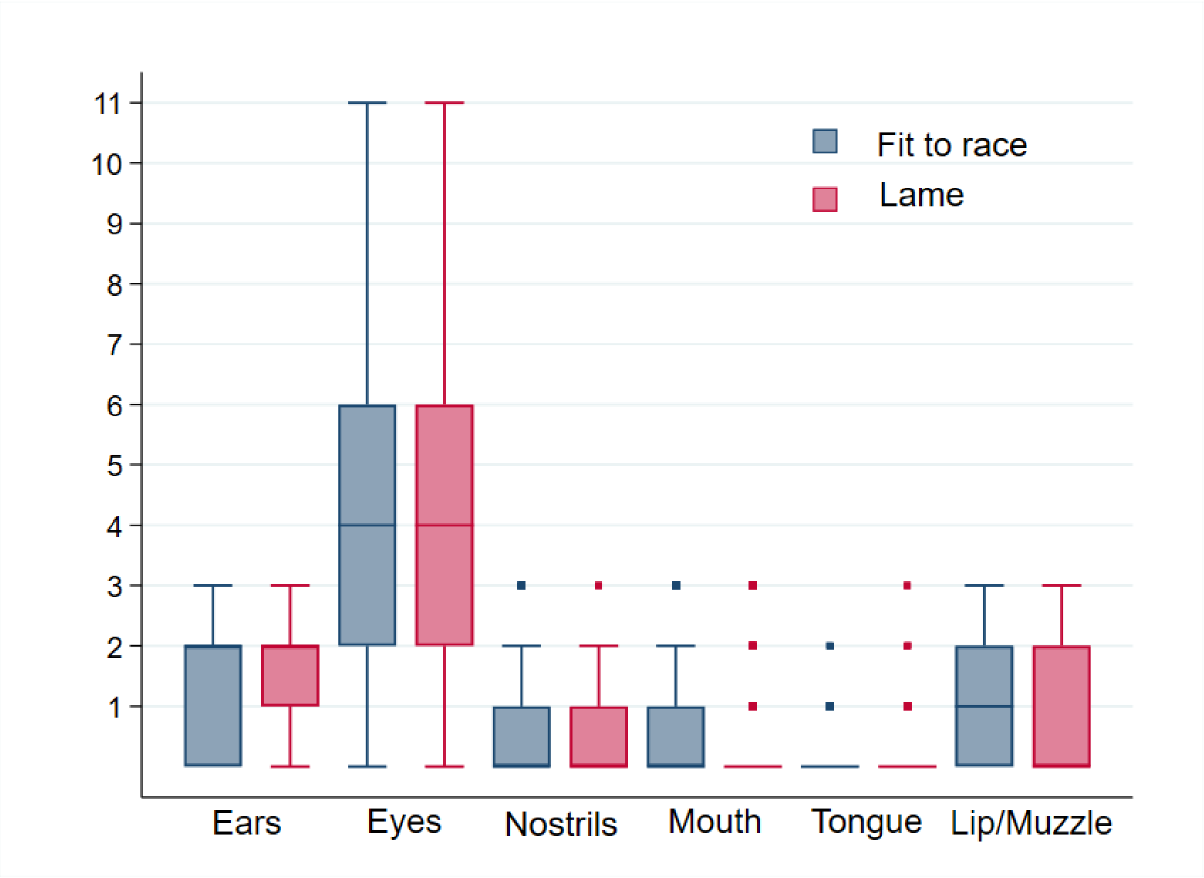
Table 3:4 Pain scores allocated to predictor by Dyson et al.

	Descriptor	Definition	Pain score
Head position	Twisted – nose to left or right	Head tipped to one side of the vertical axis	1
Ears	Ears forwards	Both ears forward with pinnae facing forwards	0
	Ears erect and parallel	Both ears vertical with pinnae facing forwards	0
	Both ears erect & to side	Both ears erect and pinnae point to side (divergent) i.e., 180^{0} different directions to the side	2

	One ear forward & one to side	Both ears erect one to the front, and one to the side with pinna pointing to side, 90^o different directions	1
	One ear to side & one back	One ear to the side (divergent) and one pinned back - 90^o different directions	2
	One ear forward & one back	One ear erect and facing forwards and one pinned backwards	1
	Both ears back	Both ears erect pinned back towards the neck	3
Eyes	Orbital tightening not present	Eye lids are well apart, not closed	0
	Orbital tightening moderately present	Eye lids are partially closed	1
	Orbital tightening obviously present	Eye lids are almost completely closed	3
	Eye white showing	Sclera exposed	3
	Eye round shape, relaxed	Round-shaped eye	0
	Eye almond shape	Almond/ diamond-shaped eye	1
	Tension above the eye	Tension of m. levator anguli oculi medialis	2
	Tension caudal to the eye	Contraction of muscles making zygomatic arch more easily visible	2
	Intense stare	Glazed look	2
Nostrils	Relaxed, neutral	Nostril cartilage in neutral position, tear-drop shape	0
	Open wide	Nostril cartilage lifted – mediolateral widening, circular in shape	0
	Tense	Lateral rim of nostril pulled back or down towards the lips, angled sides	1
	Wrinkle between nostrils	Wrinkle between nostrils	1
	Wrinkle between nostril and lip	Wrinkle folds ventral to nostril towards lip	1
Mouth	Relaxed and neutral	Lips closed	0
	Lips slightly separated	Lips separated, cannot see teeth	1
	Lips separated some teeth visible	Lips separated can see some teeth, but **no gum**, and **teeth are not apart**	1
	Lips separated some teeth visible	Lips separated can see some teeth, but **no gum**, and **teeth are apart**	1
	Lips separated and teeth visible but apposed	Lips open showing teeth **and gum** and teeth which are apposed	2
	Mouth open, teeth exposed and separated, but no tongue	Mouth open i.e., teeth slightly separated, but cannot see tongue	2
	Mouth open, teeth exposed and widely separated, but tongue not visible	Mouth open i.e., teeth widely separated, but cannot see tongue	2
	Mouth open, teeth exposed and separated and tongue visible	Mouth open i.e., teeth **widely** separated, exposing tongue	3
	Mouth open, teeth exposed and separated and tongue visible	Mouth open i.e., teeth **slightly** separated, exposing tongue	2
Upper lip	Relaxed	Muzzle relaxed with curved contour in line with lower lip	0
	Muzzle tense	Muzzle tense and angled	1
	Muzzle tense and upper muzzle extended	Muzzle tense and angled and upper muzzle extended	2
Lower lip	Relaxed	Muzzle relaxed with curved contour	0
	Muzzle tense	Muzzle tense and angled	1

Tongue	Tongue in	Tongue not seen	0
	Tongue seen inside oral cavity	Tongue seen inside oral cavity	1
	Tip of tongue protruding, no teeth	Tip of tongue protruding, cannot see teeth	1
	Tip of tongue protruding, teeth	Tip of tongue protruding, can see teeth	2
	Large part of tongue out, but teeth not exposed	Large part of tongue out, but teeth not exposed	2
	Large part of tongue out and teeth exposed	Large part of tongue out and teeth exposed	3

Table 3:5 Boxplot of facial regions score, stratified by lameness status (lame or fit to race)



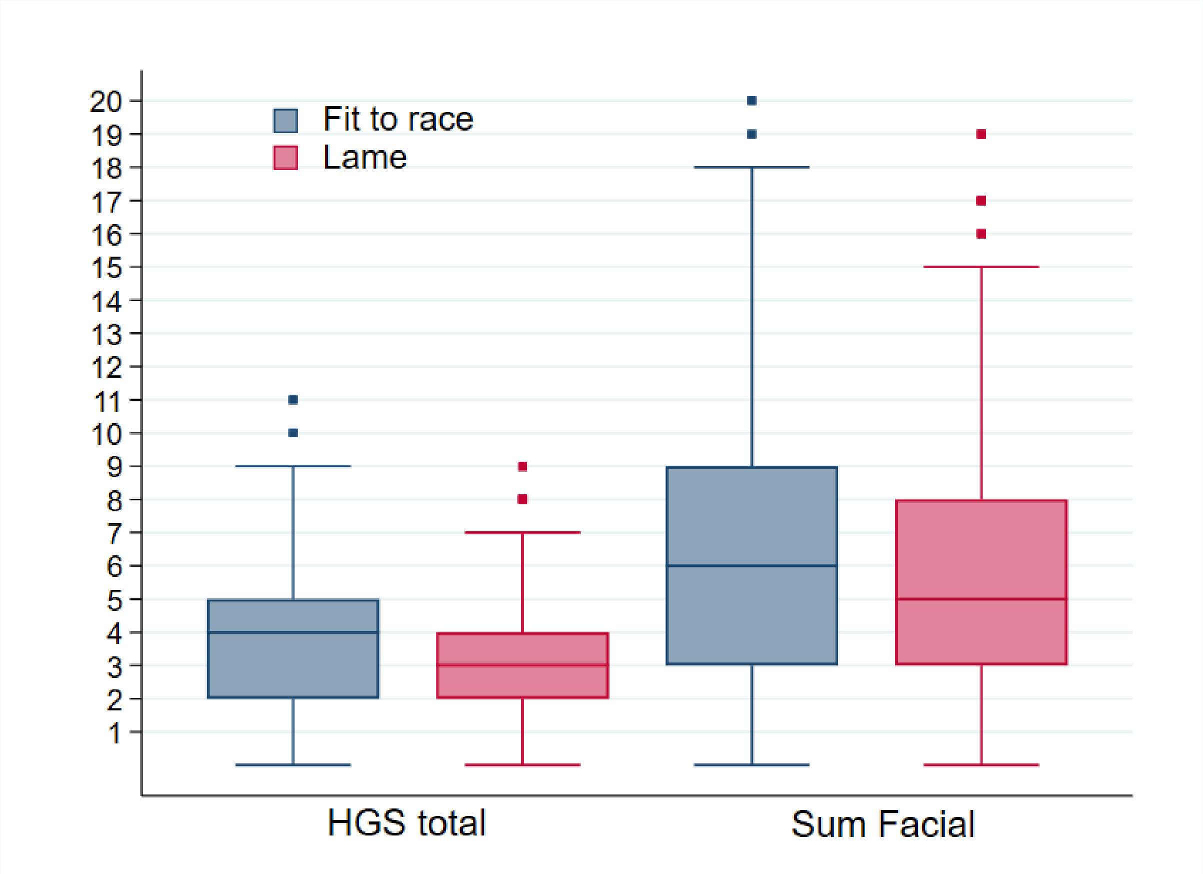


Table 3:6 The interobserver reliability between observer 1 and observer 2 for the first observation and second observation of n=38 Thoroughbred racehorses

	Time point 1			Time Point 2		
Behaviour Parameter	Agreement %	Observations not scored (/740)	Inter reliability (κ)	Agreement %	Observations not scored (/760)	Inter reliability (κ)
Stiffly ears	88.5	15	0.55 (0.48,0.62)	86.2	4	0.50 (0.43,0.57)
Orbital Tightening	93.1	56	0.66 (0.56,0.75)	95.7	45	0.80 (0.74,0.87)
Tension Above the Eye	82.0	56	0.38 (0.30,0.46)	85.3	63	0.46 (0.38,0.54)
Prominent chewing muscles	90.0	211	0.54 (0.40,0.68)	89.8	211	0.47 (0.33,0.60)
Strained Nostrils	73.9	60	0.11 (0.04,0.19)	81.6	83	0.26 (0.17,0.34)
Mouth Strained	83.6	92	0.49(0.40,0.58)	89.5	92	0.65 (0.58,0.73)
Overall Pain	80.0	19	0.04 (-0.04,0.12)	88.6	1	0.41 (0.33,0.48)
Ears (position)	53.5	18	0.41 (0.35,0.48)	50.1	5	0.39 (0.33,0.44)

Stare (normal or intense)	62.6	68	0.23 (0.13,0.33)	66.2	81	0.29 (0.19,0.40)
Eye Tension	79.9	64	0.52 (0.41,0.63)	86.9	45	0.70 (0.62,0.78)
Sclera (exposed or not exposed)	96.3	11	0.67 (0.51,0.84)	98.4	4	0.81 (0.66,0.96)
Eye shape (Almond or Round)	60.3	73	0.26 (0.17,0.36)	55.0	47	0.21 (0.13,0.28)
Tension caudal to eye	74.9	54	0.49 (0.40,0.59)	78.5	46	0.56 (0.48,0.65)
Tension above eye	70.8	66	0.42 (0.32,0.53)	69.6	64	0.41 (0.32,0.50)
Nostril Tension	48.2	72	0.24 (0.17,0.30)	65.4	82	0.41 (0.33,0.49)
Wrinkle Between nostrils	67.4	245	0.24 (0.11,0.38)	66.5	327	0.22 (0.10,0.33)
Wrinkles Ventral to Nostrils	84.5	50	0.42 (0.29,0.56)	84.4	50	0.31 (0.18,0.45)
Upper Lip	54.4	84	0.24 (0.15,0.33)	73.7	55	0.45 (0.36,0.55)
Lower lip	75.0	126	0.50 (0.39,0.61)	78.9	102	0.58 (0.48,0.68)
Nose Twisted	94.5	16	0.10 (-0.06,0.27)	95.5	4	0.27 (0.03,0.50)
Mouth	80.4	59	0.54 (0.46,0.63)	78.7	79	0.53 (0.45,0.61)
Tongue	92.8	31	0.59 (0.46,0.74)	91.8	40	0.56 (0.43,0.70)
Total		1546			1530	

Table 3:7 Association between the HGS and FEReq, and Lameness Locator measures on a continuous scale

		HMIN			**PMIN**			**HMAX**			**PMAX**		
PREDICTOR	**N**	**COEF.**	95% CI	P-VALUE	**COEF.**	95% CI	P-VALUE	**COEF.**	95% CI	P-VALUE	**COEF.**	95% CI	P-VALUE
Stiffly backward ears				**0.02**									
Not present	524	ref			ref			ref			ref		
Moderately present	772	0.71	-1.87,3.30	0.580	0.21	-0.62,1.04	0.607	0.38	-0.93,1.69	0.561	-0.75	-1.57,0.08	0.076
Obviously present	185	-3.40	-7.95,1.15	0.139	1.40	-1.32,4.11	0.304	-0.15	-2.97,2.67	0.917	-1.02	-3.22,1.17	0.352
Orbital Tightening													
Not present	1013	ref			ref			ref			ref		
Moderately present	289	-2.36	-6.58,1.86	0.265	0.35	-1.22,1.91	0.655	-1.81	-3.58,-0.47	**0.044**	0.05	-0.84,0.94	0.910
Obviously present	97	-1.45	-5.88,2.98	0.511	-0.28	-2.71,2.15	0.817	-1.66	-4.51,1.19	0.245	-0.33	-1.87,1.21	0.664
Tension above the eye													
Not present	483	ref			ref			ref			ref		
Moderately present	682	-1.67	-5.76,2.42	0.413	-0.05	-1.41,1.30	0.936	1.96	0.02,3.90	**0.048**	-0.37	-2.16,1.41	0.675
Obviously present	216	-3.16	-9.46,3.15	0.317	0.15	-1.97,2.27	0.888	2.28	-0.98,5.54	0.165	-0.08	-2.27,2.12	0.945
****Prominent Chewing Muscles**													**0.020**
Not present	638	ref			ref			ref			ref		
Moderately present	384	-4.52	-8.47,-0.56	**0.027**	1.04	-0.09,2.16	0.069	-0.16	-2.03,1.72	0.865	1.28	0.27,2.30	**0.014**
Obviously present	56	-4.42	-8.21,-0.53	**0.027**	0.66	-1.09,2.41	0.447	-0.37	-2.82,2.07	0.758	0.24	-0.94,1.42	0.681
***Strained Nostrils**													
Not present	505	ref			ref			ref			ref		
Moderately present	719	-0.17	-3.12,2.78	0.907	-0.16	-1.03,0.72	0.721	0.66	-0.70,2.02	0.331	-0.67	-1.70,0.37	0.201
Obviously present	133	-2.92	-6.92,1.08	0.148	-0.56	-2.07,0.95	0.457	-0.46	-2.68,1.76	0.676	-0.32	-1.94,1.30	0.689
****Mouth Strained**													
Not present	552	ref			ref			ref			ref		
Moderately present	519	-1.74	-6.09,2.62	0.425	-0.66	-2.04,0.71	0.335	0.02	-1.65,1.67	0.989	-1.21	-3.21,0.80	0.231

Obviously present	245	-2.10	-8.36,4.16	0.501	-0.71	-3.02,1.60	0.538	0.92	-2.12,3.95	0.541	-1.46	-3.84,0.92	0.221
***Overall Pain**													
No Pain	753	ref			ref			ref			ref		
Mild Pain	472	-2.13	-4.09,-0.17	**0.034**	-0.15	-0.85,0.54	0.663	0.18	-0.83,1.19	0.723	-0.57	-1.81,0.67	0.356
Moderate Pain	192	-0.93	-4.22,2.36	0.569	-0.12	-1.48,1.23	0.856	1.00	-0.90,2.88	0.293	-0.76	-1.98,0.45	0.212
Severe Pain	63	-3.17	-8.15,1.82	0.206	0.04	-2.11,0.42	0.968	1.01	-1.45,3.49	0.408	-1.24	-3.18,0.71	0.206
Ear Position													
Ears Forward	403	ref			ref			ref			ref		
Erect to side	477	1.06	-2.38,4.51	0.537	0.04	-1.05,1.13	0.936	0.60	-1.41,2.62	0.548	-1.16	-2.15	0.024
One ear forward, one ear back	114	1.15	-2.10,4.41	0.476	-0.04	-0.90,0.81	0.919	0.28	-1.43,1.98	0.744	-0.20	-1.16	0.676
One ear forward, one erect to side	128	2.09	-2.08,6.26	0.317	0.73	-0.45,1.91	0.220	0.74	-1.12,2.60	0.425	0.88	-1.33	0.425
One to side, one to back	154	-1.45	-6.60,3.70	0.571	1.00	-0.70,2.70	0.241	0.04	-2.17,2.26	0.966	-1.08	-2.72	0.188
Both ears back	183	-3.22	-8.10,1.65	0.188	1.37	-1.23,4.33	0.266	-0.06	-3.05,2.92	0.964	-1.07	-3.21	0.319
***Stare**													
Normal	859	ref			ref			ref			ref		
Stare	492	-0.63	-2.79,1.53	0.557	0.07	-0.76,0.91	0.859	0.79	-0.53,2.11	0.233	-0.25	-1.00,0.50	0.496
Eye Tension													
Not present	999	ref			ref			ref			ref		
Moderately present	282	-2.35	-6.61,1.91	0.271	0.40	-1.21,2.01	0.619	-1.73	-3.50,0.03	**0.054**	0.14	-0.81,1.08	0.771
Obviously present	90	-1.41	-5.87,3.06	0.527	-0.42	-2.81,1.97	0.723	-1.60	-4.47,1.26	0.262	-0.45	-2.00,1.09	0.556
Sclera													
Not exposed	1395	ref			ref			ref			ref		
Exposed	77	8.99	-2.69,20.68	0.127	0.21	-0.95,1.37	0.883	2.43	-2.16,7.03	0.289	-1.00	-2.48,-.48	0.179
***Eye Shape**													
Round	715	ref			ref			ref			ref		
Almond	665	-1.01	-3.51,1.50	0.422	0.21	-0.95,1.37	0.715	-1.31	-2.84,0.23	0.093	0.43	-0.51,1.37	0.358

Tension caudal to eye													
Not present	590	ref			ref			ref			ref		
Present	791	-3.24	-8.08,1.59	0.182	0.44	-0.92,1.81	0.514	0.74	-1.57,3.04	0.522	0.68	-0.60,1.95	0.291
***Tension above eye**													
Not present	662	ref			ref			ref			ref		
Present	708	-0.55	-4.6,3.47	0.782	-0.38	-1.61,0.85	0.536	2.19	0.18,4.20	**0.034**	-0.46	-2.04,1.12	0.559
*****Strained Nostrils**													
Relaxed	477	ref			ref			ref			ref		
Circular	704	-0.35	-3.55,2.86	0.828	-0.10	-1.05,0.84	0.827	0.72	-0.57,2.01	0.268	-0.79	-2.12,0.53	0.231
Tense	164	-3.35	-7.35,0.65	0.098	-0.40	-1.84,1.04	0.575	-0.63	-2.80,1.65	0.561	-0.30	-1.78,1.17	0.678
*****Wrinkles between nostrils**													
Not present	717	ref			ref			ref			ref		
Present	211	-2.80	-8.00,2.40	0.282	-0.49	-1.35,0.37	0.251	0.26	-1.60,2.12	0.775	-0.20	-0.93,0.54	0.593
***Wrinkles Ventral to Nostrils**													
Not present	1187	ref			ref			ref			ref		
Present	198	-3.78	-7.23,-0.31	**0.034**	0.06	-1.20,1.32	0.924	-0.46	-2.54,1.61	0.652	0.25	-1.20,1.71	0.729
***Upper Lip**													
Relaxed	816	ref			ref			ref			ref		
Tense	379	-2.77	-6.53,1.00	0.145	-0.24	-1.57,1.10	0.721	-0.56	-2.26,1.14	0.511	-0.52	-1.75,0.70	0.393
Tense and extended	152	-1.02	-4.13,2.08	0.508	-0.94	-2.57,0.58	0.217	-0.14	-2.34,2.08	0.900	-0.55	-1.78,0.67	0.364
****Lower Lip**													
Relaxed	609	ref			ref			ref			ref		
Tense	663	-1.95	-7.03,3.12	0.440	-0.45	-1.94,1.04	0.545	0.14	-1.88,2.16	0.889	-1.12	-3.07,0.82	0.248
***Nose Twisted**													
Not twisted	1443	ref			ref			ref			ref		
Twisted	47	8.39	-1.97,18.75	0.109	-2.51	-3.84,-1.19	**0.00**	3.57	0.05,7.08	**0.047**	-0.87	-2.58,0.85	0.312

Mouth													**0.0006**
Closed	991	ref			ref			ref			ref		
Lips slightly separated: no teeth	183	-2.63	-7.46,2.18	0.275	-0.37	-1.94,1.20	0.635	0.25	-1.88,2.39	0.812	-1.30	-2.77,0.17	0.082
Lips separated, teeth: no gum	50	-2.61	-10.26,5.03	0.493	0.09	-3.09,23.27	0.953	1.70	-1.7,5.14	0.324	-1.98	-3.90,-0.05	**0.045**
Lips separated, teeth and gum exposed	12	-2.30	-7.87,3.26	0.407	0.38	-4.27,5.02	0.871	5.3	1.29,9.26	**0.011**	-2.94	-5.21,-0.66	**0.013**
Teeth slightly separated: tongue not visible	21	0.98	-5.94,7.90	0.777	1.42	-3.03,5.86	0.523	3.71	0.27,7.16	**0.036**	-1.36	-3.14,0.43	0.134
Teeth slightly separated: tongue visible	56	-2.89	-8.31,2.52	0.286	-0.24	-1.94,4.46	0.777	0.81	-1.94,3.56	0.556	-1.18	-3.13,0.76	0.225
Mouth widely separated but cannot see tongue	9	-1.91	-9.08,5.27	0.593	0.67	-4.72,6.06	0.803	5.32	0.81,9.81	**0.022**	-2.38	-4.72,-0.04	**0.046**
Teeth widely separated: tongue visible	24	-4.10	-9.15,0.95	0.108	-0.37	-3.50,2.75	0.809	3.11	-2.63,8.85	0.280	-1.65	-4.99,1.70	0.325
Tongue				**0.0005**			**0.0000**			**0.0013**			**0.000**
Tongue in, not visible	1273	ref			ref			ref			ref		
Tongue seen inside oral cavity	103	-3.11	-7.43,1.21	0.153	-0.38	-1.90,1.14	0.616	0.26	-2.52,3.04	0.849	-1.33	-3.04,0.39	0.125
Tip of tongue protruding, no teeth	24	-6.19	-10.76,-1.62	**0.009**	0.85	-3.03,4.73	0.661	-1.09	-4.56,2.38	0.529	-0.22	-4.17,3.72	0.910
Tip of tongue protruding, can see teeth	10	-8.63	-13.02,-4.25	0.000	1.72	-1.76,5.19	0.324	-0.75	-6.38,4.88	0.789	1.09	-2.09,4.27	0.491
Large part of tongue out, but teeth not exposed	0												
Large part of tongue out and teeth exposed	1	-8.87	-12.75,-5.00	0.000	3.90	3.17,5.63	0.000	-4.11	-5.93,-2.29	0.00	2.73	1.39,4.07	0.000

*inter observer reliability of <0.40; ** parameters with >10% missing; ***parameters with >10% missing and inter-reliability of <0.40

Chapter 4: Discussion

4.1 Major findings and implications:

The main purpose of this thesis was to investigate methods of monitoring racehorse behaviour that may reflect orthopaedic pain and assist in the early detection of horses at risk of injury. The thesis investigated the use of two different methods of identifying discomfort. The inertial measurement units (IMUs) were able to monitor horses in stables with excellent sensitivity for identifying recumbency, standing and step, and good sensitivity for pawing ground and weight shifting. In contrast, the use of facial pain scores appeared unreliable at discriminating between lame racehorses and those considered "fit to race". Although this thesis did not investigate associations between behaviour and orthopaedic pain, it has laid the foundation for longitudinal studies investigating changes in behaviour and their correlation with the onset of injury.

The validation of IMUs to effectively monitor stabled racehorses could provide an objective tool for trainers to identify abnormal behaviour for individual horses. Although it may not be a substitute for good horsemanship, an IMU can monitor horses 24 hours a day using an objective measure that is repeatable whereas human monitoring is at best intermittent and becoming more challenging as stable size increases. Observations of horses over long periods of time are required to avoid missing intermittent signs, therefore automated systems are needed. Moreover, recent research has shown horses in pain hide their discomfort in the presence of people indicating that remote observation methods have advantages [119].

Awareness of being observed could also contribute to a lack of differences in facial pain scores between groups during lameness examinations. Performing lameness examinations in the environment of the racing stables poses challenges, with Thoroughbred racehorses reacting to external factors such as their environment or stress [175]. For these reasons, facial scoring may be more appropriate to use in stabled horses monitored remotely. Additionally, monitoring the live animal continuously may be more sensitive to facial pain grimace than photographs.

Prevention and early identification of orthopaedic problems relies on subjective evaluation, often associated with poor reliability between observers [176]. The high percentage of horses performing with mild gait asymmetries indicates the difficulty trainers and riders have identifying lameness [79, 99, 100] and that these horses can continue to perform in spite of gait asymmetry underscores the challenges in interpreting their significance. Horses considered fit to race had measurable gait asymmetry above the threshold of horses identified to be ‘sound’ (i.e. with a symmetric gait) [78, 80]. Using displacement values that denote ‘sound’ rather than the more lenient, ‘fit to race’, within our population of horses only 8% would have been classified as sound. Therefore, it is likely a majority of horses race with a degree of gait asymmetry. Identification of subtle gait changes that may signify impending fracture is challenging, if indeed they occur consistently at all. The relatively low incidence of catastrophic fractures makes determining when a horse is at risk even more difficult. Whilst veterinarians attempt to identify horses at an increased risk of injury [177], only 1.6% of horses identified on pre-race examinations develop an injury in the following race [178].

4.2 Recommendations for future research

Despite validation that the IMU technology used in this study is capable of classifying horse movements with sensitivity, to my knowledge there are currently no studies that have investigated the value of monitoring horse behaviour longitudinally. In dairy cattle, lying time monitored using accelerometers (such as IceTag), have shown predictive value in identifying disease [179]. Further research should focus on identifying whether certain patterns or changes in behaviour precede bone fatigue injuries. Longitudinal studies would be required to investigate the significance of behavioural measures to predict musculoskeletal injury. While monitoring horse behaviour over time, accurate records of horse health, injuries, biomarkers, and racing and training history would ideally need to be collected to develop a comprehensive risk-based prediction model to minimise subsequent injury.

Implementation of IMUs into routine stabling would require the use of a sensor that is more affordable, with longer battery life and easily attached to a stable wrap. An automated monitoring system would need to be developed to signal abnormal behaviour from baseline for individual horses. The gathering of data should be easily stored, accessed and used to interpret changes in behaviour over a period.

Other less intensive methods could also be explored (e.g. marker-less motion capture). An affordable technique requiring two charge-couple device (CCD) webcams, have been used to reconstruct three-dimensional movement with excellent accuracy [180]. This method would require a single installation of the cameras, per stable, thus favouring the fast-paced environment of a racing stable. Additionally, deep learning algorithms could be trained to identify different facial expressions, previously utilised in humans [181]. Whether there are correlations between facial expressions and IMU monitored behaviour is yet to be determined. Behaviour is often considered nonspecific, however if subtle changes are combined, the system may be more powerful for identifying problems.

4.3 Conclusion

Over the past decade there has been a shift in stable composition, with fewer trainers, each with more horses, meaning larger training yards and greater challenge for observation of behaviours one on one [3, 120]. The goal for any horse trainer is to acquire the optimal environment for the horse to perform and reach their full athletic career. The physical health of the horse is an important factor to do this and therefore the trainer and staff must know the individual horse and address minor issues early in development. This is difficult with large numbers of horses and high staff turnover. Trainers managing a large number of horses often have stables in numerous locations, and in different parts of the country. Maintaining accurate records of horses largely relies on good communication, but with differences of opinion, the subjectivity of observation and human error this method can be flawed. Consistent and standardised monitoring would allow one to recognise subtle changes in horse behaviour from what is considered normal.

This thesis validated a method of remote monitoring of behavioural changes that may signify discomfort and found limitations in previously reported methods of evaluation of expressions of discomfort for a population of horses that would be typical of those requiring close inspection. Further work is necessary to evaluate the usefulness of these novel indicators and their relationship with progressive musculoskeletal injury over time.

Chapter 5: References

1. Bailey, C., Rose, R., Reid, S. and Hodgson, D., *Wastage in the australian thoroughbred racing industry: A survey of sydney trainers.* Australian Veterinary Journal, 1997. 75(1): p. 64-66.
2. Dyson, P.K., Jackson, B.F., Pfeiffer, D.U. and Price, J.S., *Days lost from training by two- and three-year-old thoroughbred horses: A survey of seven uk training yards.* Equine Veterinary Journal, 2008. 40(7): p. 650-657.
3. *Racing australia, ltd. Factbook 2017/2018*. 2016/2017.
4. Sun, T.C., Riggs, C.M., Cogger, N., Wright, J. and Al-Alawneh, J.I., *Noncatastrophic and catastrophic fractures in racing thoroughbreds at the hong kong jockey club.* Equine Veterinary Journal. 0(0).
5. Parkin, T.D.H., Clegg, P.D., French, N.P., Proudman, C.J., Riggs, C.M., Singer, E.R., Webbon, P.M. and Morgan, K.L., *Catastrophic fracture of the lateral condyle of the third metacarpus/metatarsus in uk racehorses – fracture descriptions and pre-existing pathology.* The Veterinary Journal, 2006. 171(1): p. 157-165.
6. Boden, L.A., Anderson, G.A., Charles, J.A., Morgan, K.L., Morton, J.M., Parkin, T.D.H., Slocombe, R.F. and Clarke, A.F., *Risk of fatality and causes of death of thoroughbred horses associated with racing in victoria, australia: 1989–2004.* Equine Veterinary Journal, 2006. 38(4): p. 312-318.
7. Stover, S.M. and Murray, A., *The california postmortem program: Leading the way.* Veterinary Clinics of North America: Equine Practice, 2008. 24(1): p. 21-36.
8. Stover, S.M., *Nomenclature, classification, and documentation of catastrophic fractures and associated preexisting injuries in racehorses.* Journal of Veterinary Diagnostic Investigation, 2017. 29(4): p. 396-404.
9. Gray, S.N., Spriet, M., Garcia, T.C., Uzal, F.A. and Stover, S.M., *Preexisting lesions associated with complete diaphyseal fractures of the third metacarpal bone in 12 thoroughbred racehorses.* Journal of Veterinary Diagnostic Investigation, 2017. 29(4): p. 437-441.
10. Bennell, K.L., Malcolm, S.A., Brukner, P.D., Green, R.M., Hopper, J.L., Wark, J.D. and Ebeling, P.R., *A 12-month prospective study of the relationship between stress fractures and bone turnover in athletes.* Calcified Tissue International, 1998. 63(1): p. 80-85.
11. Holmes, J.M., Mirams, M., Mackie, E.J. and Whitton, R.C., *Thoroughbred horses in race training have lower levels of subchondral bone remodelling in highly loaded regions of the distal metacarpus compared to horses resting from training.* The Veterinary Journal, 2014. 202(3): p. 443-447.
12. Dalla Costa, E., Minero, M., Lebelt, D., Stucke, D., Canali, E. and Leach, M.C., *Development of the horse grimace scale (hgs) as a pain assessment tool in horses undergoing routine castration.* PLOS ONE, 2014. 9(3): p. e92281.
13. Price, J., Catriona, S., Welsh, E.M. and Waran, N.K., *Preliminary evaluation of a behaviour based system for assessment of post operative pain in horses following arthroscopic surgery.* Veterinary Anaesthesia and Analgesia, 2003. 30(3): p. 124-137.
14. Pritchett, L.C., Ulibarri, C., Roberts, M.C., Schneider, R.K. and Sellon, D.C., *Identification of potential physiological and behavioral indicators of postoperative pain in horses after exploratory celiotomy for colic.* Applied Animal Behaviour Science, 2003. 80(1): p. 31-43.
15. van Loon, J.P.A.M. and Van Dierendonck, M.C., *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, 2015. 206(3): p. 356-364.
16. Dyson, S., Berger, J.M., Ellis, A.D. and Mullard, J., *Can the presence of musculoskeletal pain be determined from the facial expressions of ridden horses (fereq)?* Journal of Veterinary Behavior: Clinical Applications and Research, 2017. 19: p. 78-89.
17. Rietmann, T.R., Stauffacher, M., Bernasconi, P., Auer, J.A. and Weishaupt, M.A., *The association between heart rate, heart rate variability, endocrine and behavioural pain*

measures in horses suffering from laminitis.* Journal of Veterinary Medicine Series A, 2004. 51(5): p. 218-225.

18. Riggs, C.M., Whitehouse, G.H. and Boyde, A., *Pathology of the distal condyles of the third metacarpal and third metatarsal bones of the horse.* Equine Veterinary Journal, 1999. 31(2): p. 140-148.
19. Martig, S., Chen, W., Lee, P.V.S. and Whitton, R.C., *Bone fatigue and its implications for injuries in racehorses.* Equine Veterinary Journal, 2014. 46(4): p. 408-415.
20. Martig, S., Lee, P.V.S., Anderson, G.A. and Whitton, R.C., *Compressive fatigue life of subchondral bone of the metacarpal condyle in thoroughbred racehorses.* Bone, 2013. 57(2): p. 392-398.
21. Verheyen, K., Price, J., Lanyon, L. and Wood, J., *Exercise distance and speed affect the risk of fracture in racehorses.* Bone. 39(6): p. 1322-1330.
22. Rosanowski, S.M., Chang, Y.M., Stirk, A.J. and Verheyen, K.L.P., *Risk factors for race-day fatality, distal limb fracture and epistaxis in thoroughbreds racing on all-weather surfaces in great britain (2000 to 2013).* Preventive Veterinary Medicine, 2017. 148: p. 58-65.
23. Frost, H.M., *Wolff law and bones structural adaptations to mechanical usuage - an overview for clinician*. 1994. p. 175-188.
24. Pearson, O.M. and Lieberman, D.E., *The aging of wolff's "law": Ontogeny and responses to mechanical loading in cortical bone.* American Journal of Physical Anthropology, 2004. 125(S39): p. 63-99.
25. Oftadeh, R., Perez-Viloria, M., Villa-Camacho, J.C., Vaziri, A. and Nazarian, A., *Biomechanics and mechanobiology of trabecular bone: A review.* Journal of Biomechanical Engineering, 2015. 137(1): p. 0108021-01080215.
26. Whitton, R.C., Mirams, M., Mackie, E.J., Anderson, G.A. and Seeman, E., *Exercise-induced inhibition of remodelling is focally offset with fatigue fracture in racehorses.* Osteoporosis International, 2013. 24(7): p. 2043-2048.
27. Firth, E.C., Doube, M. and Boyde, A., *Changes in mineralised tissue at the site of origin of condylar fracture are present before athletic training in thoroughbred horses.* New Zealand Veterinary Journal, 2009. 57(5): p. 278-283.
28. Goodship, A.E. and Smith, R.K.W., *Chapter 2.2 - skeletal physiology: Responses to exercise and training*, in *Equine exercise physiology*, K.W. Hinchcliff, R.J. Geor, and A.J. Kaneps, Editors. 2008, W.B. Saunders: Edinburgh. p. 81-105.
29. Firth, E.C., Rogers, C.W., Perkins, N.R., Anderson, B.H. and Grace, N.D., *Musculoskeletal responses of 2-year-old thoroughbred horses to early training. 1. Study design, and clinical, nutritional, radiological and histological observations.* New Zealand Veterinary Journal, 2004. 52(5): p. 261-71.
30. Whitton, R.C., Trope, G.D., Ghasem-Zadeh, A., Anderson, G.A., Parkin, T.D.H., Mackie, E.J. and Seeman, E., *Third metacarpal condylar fatigue fractures in equine athletes occur within previously modelled subchondral bone.* Bone, 2010. 47(4): p. 826-831.
31. Trope, G.D., Anderson, G.A. and Whitton, R.C., *Patterns of scintigraphic uptake in the fetlock joint of thoroughbred racehorses and the effect of increased radiopharmaceutical uptake in the distal metacarpal/tarsal condyle on performance.* Equine Veterinary Journal, 2011. 43(5): p. 509-515.
32. Tull, T.M. and Bramlage, L.R., *Racing prognosis after cumulative stress-induced injury of the distal portion of the third metacarpal and third metatarsal bones in thoroughbred racehorses: 55 cases (2000–2009).* J Am Vet Med Assoc, 2011. 238(10): p. 1316-1322.
33. Bani Hassan, E., Mirams, M., Ghasem-Zadeh, A., Mackie, E.J. and Whitton, R.C., *Role of subchondral bone remodelling in collapse of the articular surface of thoroughbred racehorses with palmar osteochondral disease.* Equine Veterinary Journal, 2016. 48(2): p. 228-233.

34. Diab, S.S., Stover, S.M., Carvallo, F., Nyaoke, A.C., Moore, J., Hill, A., Arthur, R. and Uzal, F.A., *Diagnostic approach to catastrophic musculoskeletal injuries in racehorses.* Journal of Veterinary Diagnostic Investigation, 2017. 29(4): p. 405-413.
35. Barr, E.D., Pinchbeck, G.L., Clegg, P.D., Boyde, A. and Riggs, C.M., *Post mortem evaluation of palmar osteochondral disease (traumatic osteochondrosis) of the metacarpo/metatarsophalangeal joint in thoroughbred racehorses.* Equine Veterinary Journal, 2009. 41(4): p. 366-371.
36. Firth, E.C., *The response of bone, articular cartilage and tendon to exercise in the horse.* Journal of Anatomy, 2006. 208(4): p. 513-526.
37. Pinchbeck, G.L., Clegg, P.D., Boyde, A. and Riggs, C.M., *Pathological and clinical features associated with palmar/plantar osteochondral disease of the metacarpo/metatarsophalangeal joint in thoroughbred racehorses.* Equine Veterinary Journal, 2013. 45(5): p. 587-592.
38. Riggs, C.M., *Osteochondral injury and joint disease in the athletic horse.* Equine Veterinary Education, 2006. 18(2): p. 100-112.
39. Powell, S.E., *Low-field standing magnetic resonance imaging findings of the metacarpo/metatarsophalangeal joint of racing thoroughbreds with lameness localised to the region: A retrospective study of 131 horses.* Equine Veterinary Journal, 2012. 44(2): p. 169-177.
40. Tranquille, C.A., Parkin, T.D. and Murray, R.C., *Magnetic resonance imaging-detected adaptation and pathology in the distal condyles of the third metacarpus, associated with lateral condylar fracture in thoroughbred racehorses.* Equine Veterinary Journal, 2012. 44(6): p. 699-706.
41. Whitton, R.C., Ayodele, B.A., Hitchens, P.L. and Mackie, E.J., *Subchondral bone microdamage accumulation in distal metacarpus of thoroughbred racehorses.* Equine Veterinary Journal. 0(0).
42. De Haan, C.E., O'Brien, T.R. and Koblik, P.D., *A radiographic investigation of third carpal bone injury in 42 racing thoroughbreds* Veterinary Radiology, 1987. 28(3): p. 88-92.
43. Cresswell, E.N., McDonough, S.P., Palmer, S.E., Hernandez, C.J. and Reesink, H.L., *Can quantitative computed tomography detect bone morphological changes associated with catastrophic proximal sesamoid bone fracture in thoroughbred racehorses?* Equine Veterinary Journal, 2018. 0(0).
44. Trope, G.D., Ghasem-Zadeh, A., Anderson, G.A., Mackie, E.J. and Whitton, R.C., *Can high-resolution peripheral quantitative computed tomography imaging of subchondral and cortical bone predict condylar fracture in thoroughbred racehorses?* Equine Veterinary Journal, 2015. 47(4): p. 428-432.
45. Tranquille, C.A., Murray, R.C. and Parkin, T.D.H., *Can we use subchondral bone thickness on high-field magnetic resonance images to identify thoroughbred racehorses at risk of catastrophic lateral condylar fracture?* Equine Veterinary Journal, 2017. 49(2): p. 167-171.
46. Alejandro Valdes-Martinez, R.D.P., *Radiology*, in *Adams & stashaks lameness in horses*, G.M. Baxter, Editor. 2011.
47. Butler, J.A., Colles, C.M., Dyson, S.J., Kold, S.E. and Poulos, P.W., *Clinical radiology of the horse*. 2017, Hoboken, United Kingdom: John Wiley & Sons, Incorporated.
48. Baldwin, C.M., Smith, M.R.W., Allen, S. and Wright, I.M., *Radiographic and arthroscopic features of third carpal bone slab fractures and their impact on racing performance following arthroscopic repair in a population of racing thoroughbreds in the uk.* Equine Veterinary Journal, 2019.
49. Salem, M., El-Shafaey, E., Mosbah, E. and Zaghloul, A., *Ultrasonographic, computed tomographic, and magnetic resonance imaging of the normal donkeys (equus asinus) digit.* Journal of Equine Veterinary Science, 2019. 74: p. 68-83.
50. Baxter, G.M., *Adams and stashak's lameness in horses sixth edition* Vol. 23. 2011. 377-377.

51. Crijns, C.P., Martens, A., Bergman, H.J., Veen, H., Duchateau, L., Bree, H.J.J. and Gielen, I.M.V.L., *Intramodality and intermodality agreement in radiography and computed tomography of equine distal limb fractures.* Equine Veterinary Journal, 2014. 46(1): p. 92-96.
52. Morgan, J.W., Santschi, E.M., Zekas, L.J., Scollay-Ward, M.C., Markel, M.D., Radtke, C.L., Sample, S.J., Keuler, N.S. and Muir, P., *Comparison of radiography and computed tomography to evaluate metacarpo/metatarsophalangeal joint pathology of paired limbs of thoroughbred racehorses with severe condylar fracture.* Veterinary Surgery, 2006. 35(7): p. 611-617.
53. Desbrosse, F.G., Vandeweerd, J.-M.E.F., Perrin, R.A.R., Clegg, P.D., Launois, M.T., Brogniez, L. and Gehin, S.P., *A technique for computed tomography (ct) of the foot in the standing horse.* Equine Veterinary Education, 2008. 20(2): p. 93-98.
54. Bogers, S.H., Rogers, C.W., Bolwell, C.F., Roe, W.D., Gee, E.K. and McIlwraith, C.W., *Impact of race training on volumetric bone mineral density and its spatial distribution in the distal epiphysis of the third metatarsal bone of 2-year-old horses.* The Veterinary Journal, 2014. 201(3): p. 353-358.
55. Yamada, K., Sato, F., Higuchi, T., Nishihara, K., Kayano, M., Sasaki, N. and Nambo, Y., *Experimental investigation of bone mineral density in thoroughbreds using quantitative computed tomography.* Journal of Equine Science, 2015. 26(3): p. 81-87.
56. Genton, M., Vila, T., Olive, J. and Rossignol, F., *Standing mri for surgical planning of equine fracture repair.* Veterinary Surgery, 2019. 48(8): p. 1372-1381.
57. Hoskinson, J.J., *Equine nuclear scintigraphy: Indications, uses, and techniques.* Veterinary Clinics of North America: Equine Practice, 2001. 17(1): p. 63-74.
58. Dyson, S.J., *Equine scintigraphy*. 2003, Newmarket: Equine Veterinary Journal.
59. Smith, M. and Smith, R., *Diagnostic ultrasound of the limb joints, muscle and bone in horses.* In Practice, 2008. 30(3): p. 152-159.
60. Frisbie, D.D., Mc Ilwraith, C.W., Arthur, R.M., Blea, J., Baker, V.A. and Billinghurst, R.C., *Serum biomarker levels for musculoskeletal disease in two- and three-year-old racing thoroughbred horses: A prospective study of 130 horses.* Equine Veterinary Journal, 2010. 42(7): p. 643-651.
61. Frisbie, D.D., Duffy, E., Arthur, R., Blea, J., Baker, V., Billinghurst, R.C. and McIlwraith, C.W., *Prospective clinical study assessing serum biomarkers for musculoskeletal disease in 2-3-year-old racing thoroughbreds.* Havemeyer Foundation Monograph Series, 2008(No.22): p. 11-12.
62. Greenwood, C., Clement, J., Dicken, A., Evans, J.P.O., Lyburn, I., Martin, R.M., Rogers, K., Stone, N. and Zioupos, P., *Towards new material biomarkers for fracture risk.* Bone, 2016. 93: p. 55-63.
63. Wayne, M.C., E., K.C., D., F.D., B., L.C., D., C.P., J., P.M., A., K.M., Stina, E., Sheila, L., A., S.R., J., S.L., S., L.L., and B., K.V., *Biomarkers for equine joint injury and osteoarthritis.* Journal of Orthopaedic Research, 2018. 36(3): p. 823-831.
64. Frisbie, D.D., McIlwraith, C.W. and de Grauw, J.C., *10 - synovial fluid and serum biomarkers*, in *Joint disease in the horse (second edition)*, C.W. McIlwraith, et al., Editors. 2016, W.B. Saunders: Edinburgh. p. 179-191.
65. Price, J.S., Jackson, B., Eastell, R., Wilson, A.M., Russell, R.G.G., Lanyon, L.E. and Goodship, A.E., *The response of the skeleton to physical training: A biochemical study in horses.* Bone, 1995. 17(3): p. 221-227.
66. Hiney, K.M., Potter, G.D., Gibbs, P.G. and Bloomfield, S.M., *Response of serum biochemical markers of bone metabolism totraining in the juvenile racehorse.* Journal of Equine Veterinary Science, 2000. 20(12): p. 851-857.
67. Bussières, G., Jacques, C., Lainay, O., Beauchamp, G., Leblond, A., Cadoré, J.L., Desmaizières, L.M., Cuvelliez, S.G. and Troncy, E., *Development of a composite orthopaedic pain scale in horses.* Research in Veterinary Science, 2008. 85(2): p. 294-306.

68. Lindegaard, C., Thomsen, M.H., Larsen, S. and Andersen, P.H., *Analgesic efficacy of intra-articular morphine in experimentally induced radiocarpal synovitis in horses.* Veterinary Anaesthesia and Analgesia, 2010. 37(2): p. 171-185.
69. Gleerup, K.B. and Lindegaard, C., *Recognition and quantification of pain in horses: A tutorial review.* Equine Veterinary Education, 2016. 28(1): p. 47-57.
70. Raekallio, M., Taylor, P.M. and Bennett, R.C., *Preliminary investigations of pain and analgesia assessment in horses administered phenylbutazone or placebo after arthroscopic surgery.* Veterinary Surgery, 1997. 26(2): p. 150-155.
71. Davidson, E.J., *Lameness evaluation of the athletic horse.* Veterinary Clinics of North America: Equine Practice, 2018. 34(2): p. 181-191.
72. Jeffcott, L.B., Rossdale, P.D., Freestone, J., Frank, C.J. and Towers-Clark, P.F., *An assessment of wastage in thoroughbred racing from conception to 4 years of age.* Equine Veterinary Journal, 1982. 14(3): p. 185-198.
73. Buchner, H.H.F., Savelberg, H.H.C.M., Schamhardt, H.C. and Barneveld, A., *Head and trunk movement adaptations in horses with experimentally induced fore- or hindlimb lameness.* Equine Veterinary Journal, 1996. 28(1): p. 71-76.
74. Halling Thomsen, M., Tolver Jensen, A., Sørensen, H., Lindegaard, C. and Haubro Andersen, P., *Symmetry indices based on accelerometric data in trotting horses.* Journal of Biomechanics, 2010. 43(13): p. 2608-2612.
75. Weishaupt, M.A., Wiestner, T., Hogg, H.P., Jordan, P. and Auer, J.A., *Vertical ground reaction force–time histories of sound warmblood horses trotting on a treadmill.* The Veterinary Journal, 2004. 168(3): p. 304-311.
76. Hewetson, M., Christley, R.M., Hunt, I.D. and Voute, L.C., *Investigations of the reliability of observational gait analysis for the assessment of lameness in horses.* Veterinary Record, 2006. 158(25): p. 852-858.
77. Pfau, T., Robilliard, J.J., Weller, R., Jespers, K., Eliashar, E. and Wilson, A.M., *Assessment of mild hindlimb lameness during over ground locomotion using linear discriminant analysis of inertial sensor data.* Equine Veterinary Journal, 2007. 39(5): p. 407-413.
78. Pfau, T., Sepulveda Caviedes, M.F., McCarthy, R., Cheetham, L., Forbes, B. and Rhodin, M., *Comparison of visual lameness scores to gait asymmetry in racing thoroughbreds during trot in-hand.* Equine Veterinary Education, 2018. 0(0).
79. Pfau, T., Jennings, C., Mitchell, H., Olsen, E., Walker, A., Egenvall, A., Tröster, S., Weller, R. and Rhodin, M., *Lungeing on hard and soft surfaces: Movement symmetry of trotting horses considered sound by their owners.* Equine Veterinary Journal, 2016. 48(1): p. 83-89.
80. Keegan, K.G., Dent, E.V., Wilson, D.A., Janicek, J., Kramer, J., Lacarrubba, A., Walsh, D.M., Cassells, M.W., Esther, T.M., Schiltz, P., Frees, K.E., Wilhite, C.L., Clark, J.M., Pollitt, C.C., Shaw, R., and Norris, T., *Assessment of repeatability of a wireless, inertial sensor–based lameness evaluation system for horses.* Equine Veterinary Journal, 2010. 42(2): p. 92-97.
81. Keegan, K., *Objective assessment of lameness.* Adams and Stashak's Lameness in Horses. Wiley-Blackwell, Chichester, West Sussex, UK, 2011: p. 154-164.
82. Uhlir, C., Licka, T., Kübber, P., Peham, C., Scheidl, M. and Girtler, D., *Compensatory movements of horses with a stance phase lameness.* Equine Veterinary Journal, 1997. 29(S23): p. 102-105.
83. Fuller, C.J., Bladon, B.M., Driver, A.J. and Barr, A.R.S., *The intra- and inter-assessor reliability of measurement of functional outcome by lameness scoring in horses.* The Veterinary Journal, 2006. 171(2): p. 281-286.
84. McCracken, M.J., Kramer, J., Keegan, K.G., Lopes, M., Wilson, D.A., Reed, S.K., LaCarrubba, A. and Rasch, M., *Comparison of an inertial sensor system of lameness quantification with subjective lameness evaluation.* Equine Veterinary Journal, 2012. 44(6): p. 652-656.
85. Bosch, S., Serra Bragança, F., Marin-Perianu, M., Marin-Perianu, R., van der Zwaag, B., Voskamp, J., Back, W., van Weeren, R. and Havinga, P., *Equimoves: A wireless networked*

inertial measurement system for objective examination of horse gait. Sensors, 2018. 18(3): p. 850.

86. Arkell, M., Archer, R.M., Guitian, F.J. and May, S.A., *Evidence of bias affecting the interpretation of the results of local anaesthetic nerve blocks when assessing lameness in horses.* Veterinary Record, 2006. 159(11): p. 346-348.
87. Pfau, T., Witte, T.H. and Wilson, A.M., *A method for deriving displacement data during cyclical movement using an inertial sensor.* Journal of Experimental Biology, 2005. 208(13): p. 2503-2514.
88. Kelmer, G., Keegan, K.G., Kramer, J., Wilson, D.A., Pai, F.P. and Singh, P., *Computer-assisted kinematic evaluation of induced compensatory movements resembling lameness in horses trotting on a treadmill.* American Journal of Veterinary Research, 2005. 66(4): p. 646-55.
89. Barr, A.R.S., *Where does it hurt?* Equine Veterinary Journal, 2012. 44(6): p. 627-628.
90. Keegan, K.G., Yonezawa, Y., Pai, P.F. and Wilson, D.A., *Accelerometer-based system for the detection of lameness in horses.* Biomed Sci Instrum, 2002. 38: p. 107-12.
91. Pfau, T., Boultbee, H., Davis, H., Walker, A. and Rhodin, M., *Agreement between two inertial sensor gait analysis systems for lameness examinations in horses.* Equine Veterinary Education, 2016. 28(4): p. 203-208.
92. Maliye, S., Voute, L., Lund, D. and Marshall, J.F., *An inertial sensor-based system can objectively assess diagnostic anaesthesia of the equine foot.* Equine Veterinary Journal, 2013. 45(S45): p. 26-30.
93. Rungsri, P.K., Staecker, W., Leelamankong, P., Estrada, R.J., Schulze, T. and Lischer, C.J., *Use of body-mounted inertial sensors to objectively evaluate the response to perineural analgesia of the distal limb and intra-articular analgesia of the distal interphalangeal joint in horses with forelimb lameness.* Journal of Equine Veterinary Science, 2014. 34(8): p. 972-977.
94. Weishaupt, M.A., Wiestner, T., Hogg, H.P., Jordan, P., Auer, J.A. and Barrey, E., *Assessment of gait irregularities in the horse: Eye vs. Gait analysis.* Equine Veterinary Journal, 2001(33): p. 135-40.
95. Ishihara, A., Bertone, A.L. and Rajala-Schultz, P.J., *Association between subjective lameness grade and kinetic gait parameters in horses with experimentally induced forelimb lameness.* American Journal of Veterinary Research, 2005. 66(10): p. 1805-1815.
96. Kai, M., Aoki, O., Hiraga, A., Oki, H. and Tokuriki, M., *Use of an instrument sandwiched between the hoof and shoe to measure vertical ground reaction forces and three-dimensional acceleration at the walk, trot, and canter in horses.* American Journal of Veterinary Research, 2000. 61(8): p. 979-85.
97. Parkes, R.S., Weller, R., Groth, A.M., May, S. and Pfau, T., *Evidence of the development of 'domain-restricted' expertise in the recognition of asymmetric motion characteristics of hindlimb lameness in the horse.* Equine Veterinary Journal, 2009. 41(2): p. 112-7.
98. Dyson, S., *Can lameness be graded reliably?* Equine Veterinary Journal, 2011. 43(4): p. 379-382.
99. Rhodin, M., Egenvall, A., Haubro Andersen, P. and Pfau, T., *Head and pelvic movement asymmetries at trot in riding horses in training and perceived as free from lameness by the owner.* PLOS ONE, 2017. 12(4): p. e0176253.
100. Pfau, T., Parkes, R.S., Burden, E.R., Bell, N., Fairhurst, H. and Witte, T.H., *Movement asymmetry in working polo horses.* Equine Veterinary Journal, 2016. 48(4): p. 517-522.
101. Dyson, S., Berger, J., Ellis, A.D. and Mullard, J., *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, 2018. 23: p. 47-57.
102. Sepulveda Caviedes, M.F., Forbes, B.S. and Pfau, T., *Repeatability of gait analysis measurements in thoroughbreds in training.* Equine Veterinary Journal, 2018. 50(4): p. 513-518.

103. Egger, C.M., Love, L. and Doherty, T., *Pain management in veterinary practice*. 2013, Somerset, UNITED STATES: Wiley.

104. Casper Lindegaard, K.B.G., and Pia Haubro Andersen, *Pathophysiology of pain*, in *The equine acute abdomen*. 2017.

105. Wagner, A.E., *Effects of stress on pain in horses and incorporating pain scales for equine practice.* Veterinary Clinics of North America: Equine Practice, 2010. 26(3): p. 481-492.

106. *Oxford dictionary of english*, in *Oxford Dictionary of English*, A. Stevenson, Editor. 2010, Oxford University Press.

107. Ashley, F.H., Waterman-Pearson, A.E. and Whay, H.R., *Behavioural assessment of pain in horses and donkeys: Application to clinical practice and future studies.* Equine Veterinary Journal, 2005. 37(6): p. 565-575.

108. Gleerup, K.B., Forkman, B., Lindegaard, C. and Andersen, P.H., *An equine pain face.* Veterinary Anaesthesia and Analgesia, 2015. 42(1): p. 103-114.

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

110. Darwin, C., *The expression of the emotions in man and animals*. 1916, New York ;: D. Appleton and Co.

111. Wathan, J., Burrows, A.M., Waller, B.M. and McComb, K., *Equifacs: The equine facial action coding system.* PLOS ONE, 2015. 10(8): p. e0131738.

112. Hadjistavropoulos, T., Chapelle, D.L., Hadjistavropoulos, H.D., Green, S. and Asmundson, G.J.G., *Using facial expressions to assess musculoskeletal pain in older persons.* European Journal of Pain, 2002. 6(3): p. 179-187.

113. 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., Matsumiya, L., Sorge, R.E., Sotocinal, S.G., Tabaka, J.M., Wong, D., van den Maagdenberg, A.M.J.M., Ferrari, M.D., Craig, K.D., and Mogil, J.S., *Coding of facial expressions of pain in the laboratory mouse.* Nature Methods, 2010. 7: p. 447.

114. Sotocina, S.G., Sorge, R.E., Zaloum, A., Tuttle, A.H., Martin, L.J., Wieskopf, J.S., Mapplebeck, J.C., Wei, P., Zhan, S., Zhang, S., McDougall, J.J., King, O.D., and Mogil, J.S., *The rat grimace scale: A partially automated method for quantifying pain in the laboratory rat via facial expressions.* Molecular Pain, 2011. 7: p. 1744-8069-7-55.

115. Keating, S.C.J., Thomas, A.A., Flecknell, P.A. and Leach, M.C., *Evaluation of emla cream for preventing pain during tattooing of rabbits: Changes in physiological, behavioural and facial expression responses.* PLOS ONE, 2012. 7(9): p. e44437.

116. Dalla Costa, E., Stucke, D., Dai, F., Minero, M., Leach, M.C. and Lebelt, D., *Using the horse grimace scale (hgs) to assess pain associated with acute laminitis in horses (equus caballus).* Animals, 2016. 6(8): p. 47.

117. Mullard, J., Berger, J.M., Ellis, A.D. and Dyson, S., *Development of an ethogram to describe facial expressions in ridden horses (fereq).* Journal of Veterinary Behavior, 2017. 18: p. 7-12.

118. Dyson, S., Berger, J.M., Ellis, A.D. and Mullard, J., *Behavioral observations and comparisons of nonlame horses and lame horses before and after resolution of lameness by diagnostic analgesia.* Journal of Veterinary Behavior, 2018. 26: p. 64-70.

119. Torcivia, C. and McDonnell, S., *In-person caretaker visits disrupt ongoing discomfort behavior in hospitalized equine orthopedic surgical patients.* Animals, 2020. 10(2): p. 210.

120. *Racing australia, ltd. Factbook 2007/2008.* 2007/2008.

121. Mathie, M.J., Coster, A.C.F., Lovell, N.H. and Celler, B.G., *Detection of daily physical activities using a triaxial accelerometer.* Medical and Biological Engineering and Computing, 2003. 41(3): p. 296-301.

122. Barwick, J., Lamb, D., Dobos, R., Schneider, D., Welch, M. and Trotter, M., *Predicting lameness in sheep activity using tri-axial acceleration signals.* Animals, 2018. 8(1): p. 12.
123. Razal, C.B., Bryant, J., & Miller, L. J., *Monitoring the behavioral and adrenal activity of giraffe (giraffa camelopardalis) to assess welfare during seasonal housing changes. Animal behavior and cognition,.* 4, 2017. 2: p. 154–164.
124. Bachmann, M., Wensch-Dorendorf, M., Hoffmann, G., Steinhöfel, I., Bothendorf, S. and Kemper, N., *Pedometers as supervision tools for mares in the prepartal period.* Applied Animal Behaviour Science, 2014. 151: p. 51-60.
125. Fehlmann, G., O'Riain, M.J., Hopkins, P.W., O'Sullivan, J., Holton, M.D., Shepard, E.L.C. and King, A.J., *Identification of behaviours from accelerometer data in a wild social primate.* Animal Biotelemetry, 2017. 5(1): p. 6.
126. Vianey, L.B., Theoni, P., Roland, L., A., P.T., Y., W.Y., Megan, M. and P., P.Y., *Analysis of animal accelerometer data using hidden markov models.* Methods in Ecology and Evolution, 2017. 8(2): p. 161-173.
127. Whiteside, D., Cant, O., Connolly, M. and Reid, M., *Monitoring hitting load in tennis using inertial sensors and machine learning.* Int J Sports Physiol Perform, 2017. 12(9): p. 1212-1217.
128. Shobha, G. and Rangaswamy, S., *Chapter 8 - machine learning*, in *Handbook of statistics*, V.N. Gudivada and C.R. Rao, Editors. 2018, Elsevier. p. 197-228.
129. Jebur, S.O.a.A.N.a.H.H., *Machine learning techniques for anomaly detection: An overview.* International Journal of Computer Applications, 2013. 79(2): p. 33-41.
130. Martiskainen, P., Järvinen, M., Skön, J.-P., Tiirikainen, J., Kolehmainen, M. and Mononen, J., *Cow behaviour pattern recognition using a three-dimensional accelerometer and support vector machines.* Applied Animal Behaviour Science, 2009. 119(1): p. 32-38.
131. den Uijl, I., Gómez Álvarez, C.B., Bartram, D., Dror, Y., Holland, R. and Cook, A., *External validation of a collar-mounted triaxial accelerometer for second-by-second monitoring of eight behavioural states in dogs.* PLOS ONE, 2017. 12(11): p. e0188481.
132. Kawcak, C.E., McIlwraith, C.W., Norrdin, R.W., Park, R.D. and Steyn, P.S., *Clinical effects of exercise on subchondral bone of carpal and metacarpophalangeal joints in horses.* Am J Vet Res, 2000. 61(10): p. 1252-8.
133. Lieber, B., Taylor, B.E.S., Appelboom, G., McKhann, G. and Connolly, E.S., *Motion sensors to assess and monitor medical and surgical management of parkinson disease.* World Neurosurgery, 2015. 84(2): p. 561-566.
134. Lugade, V., Fortune, E., Morrow, M. and Kaufman, K., *Validity of using tri-axial accelerometers to measure human movement - part i: Posture and movement detection.* Medical engineering & physics, 2014. 36(2): p. 169-176.
135. Mathie, M.J., Celler, B.G., Lovell, N.H. and Coster, A.C.F., *Classification of basic daily movements using a triaxial accelerometer.* Medical and Biological Engineering and Computing, 2004. 42(5): p. 679-687.
136. Aghanavesi, S., Bergquist, F., Nyholm, D., Senek, M. and Memedi, M., *Motion sensor-based assessment of parkinson's disease motor symptoms during leg agility tests: Results from levodopa challenge.* Journal of Biomedical Health and Health Informatics, 2019.
137. Elliott, K.H., Vaillant, M.L., Kato, A., Speakman, J.R. and Ropert-Coudert, Y., *Accelerometry predicts daily energy expenditure in a bird with high activity levels.* Biology Letters, 2013. 9(1): p. 20120919.
138. Graf, P.M., Wilson, R.P., Qasem, L., Hackländer, K. and Rosell, F., *The use of acceleration to code for animal behaviours; a case study in free-ranging eurasian beavers castor fiber.* PLOS ONE, 2015. 10(8): p. e0136751.
139. Gibbons, J., Medrano-Galarza, C., Marie de Passillé, A. and Rushen, J., *Lying laterality and the effect of icetag data loggers on lying behaviour of dairy cows.* Applied Animal Behaviour Science, 2012. 136(2): p. 104-107.

140. Trénel, P., Jensen, M.B., Decker, E.L. and Skjøth, F., *Quantifying and characterizing behavior in dairy calves using the icetag automatic recording device.* J Dairy Sci, 2009. 92(7): p. 3397-3401.
141. Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. , *Statistical power analyses using g*power 3.1: Tests for correlation and regression analyses.* Behavior Research Methods, 2009. 41: p. 1149-1160.
142. Ungar, E.D., Nevo, Y., Baram, H. and Arieli, A., *Evaluation of the icetag leg sensor and its derivative models to predict behaviour, using beef cattle on rangeland.* Journal of neuroscience methods, 2018. 300: p. 127-137.
143. Gabadinho, A., Ritschard, G., Müller, N.S. & Studer, M., *Analyzing and visualizing state sequences in r with traminer.* Journal of Statistical Software, 2011. 40(4): p. 1-37.
144. Chung, E.L.T., Khairuddin, N.H., Azizan, T.R.P.T. and Adamu, L., *Sleeping patterns of horses in selected local horse stables in malaysia.* Journal of Veterinary Behavior, 2018. 26: p. 1-4.
145. DuBois, C., Zakrajsek, E., Haley, D.B. and Merkies, K., *Validation of triaxial accelerometers to measure the lying behaviour of adult domestic horses.* animal, 2015. 9(1): p. 110-4.
146. Robert, B., White, B.J., Renter, D.G. and Larson, R.L., *Evaluation of three-dimensional accelerometers to monitor and classify behavior patterns in cattle.* Computers and Electronics in Agriculture, 2009. 67(1): p. 80-84.
147. Ledgerwood, D.N., Winckler, C. and Tucker, C.B., *Evaluation of data loggers, sampling intervals, and editing techniques for measuring the lying behavior of dairy cattle.* J Dairy Sci, 2010. 93(11): p. 5129-5139.
148. Eerdekens, A., Deruyck, M., Fontaine, J., Martens, L., Poorter, E.D. and Joseph, W., *Automatic equine activity detection by convolutional neural networks using accelerometer data.* Computers and Electronics in Agriculture, 2020. 168: p. 105139.
149. Thompson, C.J., Luck, L.M., Keshwani, J., Pitla, S.K. and Karr, L.K., *Location on the body of a wearable accelerometer affects accuracy of data for identifying equine gaits.* Journal of Equine Veterinary Science, 2018. 63: p. 1-7.
150. González, L.A., Bishop-Hurley, G.J., Handcock, R.N. and Crossman, C., *Behavioral classification of data from collars containing motion sensors in grazing cattle.* Computers and Electronics in Agriculture, 2015. 110: p. 91-102.
151. Hämäläinen, W.M., P.; Järvinen, M.; Skön, J.-P.; Tiirikainen, J.; Kolehmainen, M. and Mononen, J., *Computational challenges in deriving dairy cows' action patterns from accelerometer data.* Proceedings of the 22nd Nordic symposium of the International Society for Applied Ethology, 2010. Finnish Society for Applied Ethology: p. 18.
152. Smith, D., Rahman, A., Bishop-Hurley, G.J., Hills, J., Shahriar, S., Henry, D. and Rawnsley, R., *Behavior classification of cows fitted with motion collars: Decomposing multi-class classification into a set of binary problems.* Computers and Electronics in Agriculture, 2016. 131: p. 40-50.
153. White, B.J., Coetzee, J.F., Renter, D.G., Babcock, A.H., Thomson, D.U. and Andresen, D., *Evaluation of two-dimensional accelerometers to monitor behavior of beef calves after castration.* American Journal of Veterinary Research, 2008. 69(8): p. 1005-12.
154. Sweet, A.L., *Temporal discrimination by the human eye.* The American Journal of Psychology, 1953. 66: p. 185-198.
155. Dyson, S., *Musculoskeletal scintigraphy of the equine athlete.* Seminars in Nuclear Medicine, 2014. 44(1): p. 4-14.
156. Dyson, P., Jackson, BF, Pfeiffer, DU et al, *Days lost from training by two- and three-year-old thoroughbred horses: A survey of seven uk training yards.* Equine Veterinary Journal, 2008. 40(7): p. 650-657.
157. Jeffcott, L., Rossdale, PD, Freestone, J, et al, *An assessment of wastage in thoroughbred racing from conception to 4 years of age.* Equine Vet J, 1982. 14(3): p. 185-198.

158. van Weeren, P.R., Pfau, T., Rhodin, M., Roepstorff, L., Serra Bragança, F. and Weishaupt, M.A., *Do we have to redefine lameness in the era of quantitative gait analysis?* Equine Veterinary Journal, 2017. 49(5): p. 567-569.
159. Van Loon, J.P.A.M. and Van Dierendonck, M.C., *Pain assessment in horses after orthopaedic surgery and with orthopaedic trauma.* The Veterinary Journal, 2019. 246: p. 85-91.
160. Keegan, K.G., Yonezawa, Y., Pai, P.F., Wilson, D.A. and Kramer, J., *Evaluation of a sensor-based system of motion analysis for detection and quantification of forelimb and hind limb lameness in horses.* American Journal of Veterinary Research, 2004. 65(5): p. 665-70.
161. Keegan, K.G., Pai, P.F., Wilson, D.A. and Smith, B.K., *Signal decomposition method of evaluating head movement to measure induced forelimb lameness in horses trotting on a treadmill.* Equine Veterinary Journal, 2001. 33(5): p. 446-451.
162. Keegan, K.G., Kramer, J., Yonezawa, Y., Maki, H., Pai, P.F., Dent, E.V., Kellerman, T.E., Wilson, D.A. and Reed, S.K., *Assessment of repeatability of a wireless, inertial sensor-based lameness evaluation system for horses.* American Journal of Veterinary Research, 2011. 72(9): p. 1156-63.
163. Dyson, S. and Van Dijk, J., *Application of a ridden horse ethogram to video recordings of 21 horses before and after diagnostic analgesia: Reduction in behaviour scores.* Equine Veterinary Education, 2018. 0(0).
164. Friard, O., & Gamba, M. , *Boris: A free, versatile open-source event logging software for video/audio coding and live observation* Methods in Ecology and Evolution, 2016. 7(11): p. 1325-1330.
165. Tuttle, A.H., Molinaro, M.J., Jethwa, J.F., Sotocinal, S.G., Prieto, J.C., Styner, M.A., Mogil, J.S. and Zylka, M.J., *A deep neural network to assess spontaneous pain from mouse facial expressions.* Molecular Pain, 2018. 14: p. 1744806918763658.
166. Fleiss, J.L., *Statistical methods for rates and proportions*. 3rd ed. 1981, New York: Wiley.
167. Gwet, K.L., *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters*. Vol. 4th ed. 2014, Gaithersburg, MD.
168. Landis, J.R. and Koch, G.G., *The measurement of observer agreement for categorical data.* Biometrics, 1977. 33(1): p. 159-174.
169. Sim, J. and Wright, C.C., *The kappa statistic in reliability studies: Use, interpretation, and sample size requirements.* Physical Therapy, 2005. 85(3): p. 257-268.
170. Bell, R.J.W., Kingston, J.K., Mogg, T.D. and Perkins, N.R., *The prevalence of gastric ulceration in racehorses in new zealand.* New Zealand Veterinary Journal, 2007. 55(1): p. 13-18.
171. Pfau, T., Noordwijk, K., Sepulveda Caviedes, M.F., Persson-Sjodin, E., Barstow, A., Forbes, B. and Rhodin, M., *Head, withers and pelvic movement asymmetry and their relative timing in trot in racing thoroughbreds in training.* Equine Veterinary Journal, 2018. 50(1): p. 117-124.
172. Dyson, S. and Greve, L., *Subjective gait assessment of 57 sports horses in normal work: A comparison of the response to flexion tests, movement in hand, on the lunge, and ridden.* Journal of Equine Veterinary Science, 2016. 38: p. 1-7.
173. Pitts, J.B., Kramer, J., Reed, S.K., Schiltz, P., Thombs, L. and Keegan, K.G., *Effect of induced hindlimb length difference on body-mounted inertial sensor measures used to evaluate hindlimb lameness in horses.* PLOS ONE, 2020. 15(2): p. e0228872.
174. Keegan, K.G., MacAllister, C.G., Wilson, D.A., Gedon, C.A., Kramer, J., Yonezawa, Y., Maki, H. and Pai, P.F., *Comparison of an inertial sensor system with a stationary force plate for evaluation of horses with bilateral forelimb lameness.* American Journal of Veterinary Research, 2012. 73(3): p. 368-74.
175. Nogueira, G.P. and Barnabe, R.C., *Is the thoroughbred race-horse under chronic stress?* Braz J Med Biol Res, 1997. 30(10): p. 1237-9.
176. Keegan, K.G., *Reliability of equine visual lameness classification.* Veterinary Record, 2019. 184(2): p. 60-62.

177. Cohen, N.D., Peloso, J.G., Mundy, G.D., Fisher, M., Holland, R.E., Little, T.V., Misheff, M.M., Watkins, J.P., Honnas, C.M. and Moyer, W., *Racing-related factors and results of prerace physical inspection and their association with musculoskeletal injuries incurred in thoroughbreds during races.* J Am Vet Med Assoc, 1997. 211(4): p. 454-63.
178. Stover, S.M., *The epidemiology of thoroughbred racehorse injuries.* Clinical Techniques in Equine Practice, 2003. 2(4): p. 312-322.
179. Kokin. E, P.J., Veermän. I, Poikalainen. V, Vallas. M, *Icetag3dtm accelerometric device in cattle lameness detection.* Agronomy Research, 2014.
180. Rybarczyk, Y. *3d markerless motion capture: A low cost approach*. in *New Advances in Information Systems and Technologies*. 2016. Cham: Springer International Publishing.
181. Minaee, S.A., AmirAli, *Deep-emotion: Facial expression recognition using attentional convolutional network.* ArXiv, 2019.