animals
MDPI

Article

Generation of domains for the Equine Musculoskeletal Rehabilitation Outcome Score: Development by Expert Consensus

Gillian Tabor [1]*, Kathryn Nankervis[1], John Fernandes[1] and Jane Williams[1]

[1] Hartpury University;
* Correspondence: Gillian.tabor@hartpury.ac.uk

Received: date; Accepted: date; Published: date

Simple Summary: Within rehabilitation, measurements taken before, during and after treatments are used to judge patient progress and the effectiveness of prescribed treatments. To know which measurements to use for a given health conditions, practitioners must have knowledge of what should be measured, which measurement tools are available and accurate, alongside what they intend to measure. Composite outcome measures (OMs) are tools which use grouped measurement tests to monitor patient progress; they have been tested for a variety of human and canine conditions but none have been designed or tested for use in physical rehabilitation in horses. This study asked leading equine veterinarians, physiotherapists and researchers which measures should be included in an OM for use in the rehabilitation of horses. Using a process to evaluate agreement, ten areas of measurement were included in the final model: lameness, pain at rest, pain during exercise, behaviour during exercise, muscular symmetry, performance/functional capacity, behaviour at rest, palpation, balance and proprioception. Existing reliable tests used to measure these areas were evaluated and potential new measures discussed and now should be taken forward to testing as a composite outcome score to see if they are effective in measuring effectiveness of treatment.

Abstract: Outcome measures (OM) are a requirement of professional practice standards in human and canine physiotherapy practice for measurement of health status. Measures such as pain and functional capacity of specific regions are used to track treatment impact and can be used to develop optimal management strategies. To achieve comparable patient care in equine physiotherapy, OMs must be incorporated into practice, however no reliable and valid OMs exist for equine rehabilitation. This study utilised the experience and opinion of a panel of experts working in the equine rehabilitation sphere to gain consensus on the core areas (domains) to be included in a model, to lead to an OM scale for horses undergoing rehabilitation. The Delphi method and content validity ratio testing was used to determine agreement with domains reaching the critical value required for inclusion. The expert panel agreed ten domains to be included in the OM scale: lameness, pain at rest, pain during exercise, behaviour during exercise, muscular symmetry, performance/functional capacity, behaviour at rest, palpation, balance and proprioception. An OM with these domains would provide a holistic objective assessment tool which could be used by equine rehabilitation professionals in clinical practice.

Keywords: Equine; Physiotherapy; Outcome Measures; Rehabilitation; Delphi method

1. Introduction

Physiotherapy is recommended for a number of equine musculoskeletal conditions such as overriding dorsal spinous processes and thoracolumbosacral pain, soft tissue injuries such as ligament and tendon injuries and osteoarthritis [1,2]. However, the degree of detail regarding the specific physiotherapy interventions, such as for treatments including manual therapy [3], electrotherapy [1] or exercise therapy [4,5], either individually or in combination, varies between publications ranging from trials, often with low subject numbers to clinical review papers. As a result equine physiotherapists use this information in combination with their experience and clinical reasoning to select treatment interventions they consider effective [6]. For example, in rehabilitation plans for overriding dorsal spinous processes, exercises to encourage ventral flexion to separate the spinous processes [7] are used in combination with exercises to strengthen the deeper 'core' stability muscle *multifidus* [8]. For thoracolumbosacral pain, electrotherapy in the form of neuromuscular electrical stimulation [7,9] or manual therapy [10,11] are commonly applied. However due to variation in practitioners experience, and the distinct nature of each patient, there are no standardised practice guidelines for equine rehabilitation. This lack of standardisation places increased emphasis on the physiotherapists' ability to assess each horse's progress to ensure they meet their duty of care to the patient despite the current lack of an evidence base to support this decision making [12].

A common feature in published studies that include physiotherapy techniques, is a lack of objectivity when reporting on the outcome (where outcome is defined as '*any identified result arising from exposure to a causal factor or a health intervention*') [13]. Within human orthopaedic research, Chiarotto et al. [14] suggested that outcomes are inconsistently measured and reported across trials of health interventions for low back pain in humans [14]. Similarly, in equine research subjective outcomes (e.g. decisions on success based on horse-owner survey) are reported after surgery for over-riding dorsal spinous processes [7,15,16] and the treatment of sacro-iliac disease [17]. The lack of outcome measurement reduces the ability to compare findings between studies and potentially encourages selective reporting of favourable outcomes [14]. This will impact ongoing practice and *may* result in confirmation bias when assessing subsequent outcomes, thus placing the patient at risk of lack of progress, or worse still deterioration of their situation. Given the duty of care that a physiotherapist has with their patient, this remains an important issue. In human research, to reduce the heterogeneity of outcome measures (OMs) in clinical trials, there are agreed minimum sets of outcomes that should be measured by clinicians and reported for a particular health condition [14]. These specific measurement tools or techniques are known as outcome measures and a grouping of OMs can be used to form a composite outcome score that can then be used to assess the short- and long-term effect of rehabilitation for the patient [18].

OMs have been developed for use in human practice for the measurement of health status and include measures of pain and functional capacity in specific regions, used to track impact of treatment and thus the development of optimal management strategies [19]. For sport injuries, the Victorian Institute of Sport Assessment Scales for patella tendinosis and achilles tendinopathy, and the Copenhagen Hip and Groin Outcome Score are examples of OMs that have been generated to score pain, symptoms and physical function [20-22]. For dogs, outcomes can be measured with the Helsinki Chronic Pain Index, the Canine Brief Pain Index or the Finnish neurological function testing battery for dogs named the FINFUN [23-25]. These examples of composite OMs for humans and dogs have face validity, have undergone reliability and validity testing, and are used in clinical practice, however no composite OMs have been developed for equine physical rehabilitation.

To achieve comparable professional practice standards in equine physiotherapy, OMs must be incorporated into practice [6]. To date, a few equine specific OMs, that measure a single factor in clinical practice (referred to as objective markers (OBJM)), have been subject to reliability testing but there are no composite equine OMs. OBJMs include the use of pressure algometry [26,27], manual palpation scoring [28,29], posture/muscle size measurement from photographs [30], muscle dimension measurement using a flexicurve ruler [31], range of joint motion using a universal goniometer [32,33] and evaluation of pain-related behaviour [34]. Despite these studies, the use of

OBJM in clinical practice is sparse and clinicians report this being due to the lack of available, validated and reliable OBJMs [6], suggesting a lack of awareness to the available evidence.. In a recent survey, equine physiotherapists stressed that OBJMs and OMs need to be simple to use, inexpensive and relevant to the cases they see [6]. It is unknown which domains clinicians working in the equine rehabilitation industry would consider valuable to measure and how these could be combined to generate a composite outcome score specific and relevant for the cases practitioners work with. The aim of this study was to determine which domains should be measured within equine musculoskeletal rehabilitation, to develop a globally useful composite outcome score.

2. Materials and Methods

The methodology was guided by international best practice guidelines for the development of patient reported outcome measures [35] and involved iterative stages using a mixed methods approach that involved a literature review [36] and expert input. The Delphi method of gathering data was used to gain a convergence of opinion from the invited selection of veterinarians, physiotherapists and equine researchers located world-wide. The Delphi method, which is an accepted method for achieving convergence of opinion, was selected as a technique using group communication from a panel of experts [37]. Using this method, the panel members are able to review and revise their responses in the stages of the process [38] and the controlled feedback process provides anonymity to the respondents, which may be a factor in group based discussions [37].

2.1 Delphi step 1

Via email, 35 subject matter specialists, based in Europe and the United States were invited to participant in the study based on their expertise in equine rehabilitation. These included: ten equine veterinary surgeons with greater than 10 years clinical experience, all of whom are published in equine musculoskeletal health and behaviour research; fifteen UK Chartered Physiotherapists (Association of Chartered Physiotherapists in Animal Therapy, category A members) with greater than 10 years equine practice experience; and ten equine research professionals, with an interest in equine musculoskeletal rehabilitation and performance working in equine higher education institutes. Consent by participants, to be included in the Delphi process, was gained via response to the first email in step one, which also confirmed responses would be compiled anonymously.

Once invited to participate each expert was asked to reply with confirmation that they wished to be included in further rounds of the process and asked to suggest domains to be included. The term domain was defined as an area of measurement that could be included within an OM for equine musculoskeletal rehabilitation.

At this stage the number of survey rounds was not fixed and was to be determined by the degree of consensus within the panel of experts. We did, however, expect there to be between three and five rounds with the last providing a final opportunity for the experts to revise their judgments [37].

2.2 Delphi step 2

An email with a link to a questionnaire (SurveyMonkey, San Mateo, California, USA) was sent out to those experts that responded positively to being included in the Delphi panel. This stage was designed to assist selection of the domains that should be included in the final tool termed 'the equine musculoskeletal rehabilitation outcome score (TEMROS)' with the option to suggest other areas that could also be included. There were potentially a large range of domains that could be part of the outcome score, thus to keep the outcome score practitioner friendly, valid and reliable, the number of domains included needed to be limited by consensus of the Delphi panel. The experts were provided with a list of domains collated from the response of the first email round. Within the second questionnaire, each domain required the expert to mark whether the specific outcome was essential, useful but not essential, not useful or if the expert was unsure if it should be included as an area of measurement for the purpose of musculoskeletal assessment in a horse undergoing rehabilitation [39,40].

2.3 Delphi step 3

From the responses gained, a content validation process was used to agree to include or discard items listed as possible domains (Lawshe, 1975) with content validity ratio (CVR) and critical values used to confirm the level of agreement that exceeds that of chance (figure 1) [40]. Perfect agreement would result in +1 and perfect disagreement results in a CVR of -1. This process was used to identify the domains to be included in TEMROS.

$$CVR = \frac{n_e - (N/2)}{N/2}$$

Figure 1: Lawshe's (1975) content validity ratio (CVR) where n_e is the number of essential members and N is number of panel members [39].

2.4 Delphi step 4

The list of domains that met the agreement criteria were emailed to the panel of experts who were invited to comment on the final selection.

3. Results

3.1 Delphi step 1

Seven veterinary surgeons, eleven ACPAT Physiotherapists and six equine industry experts agreed to be included in the Dephi process and fifteen potential domains were suggested. These fifteen domains were taken forwards to the questionnaire in step 2.

3.2 Delphi step 2

The questionnaire was returned by 21 of the 24 experts from step 1 and the data tabulated (Figure 2)

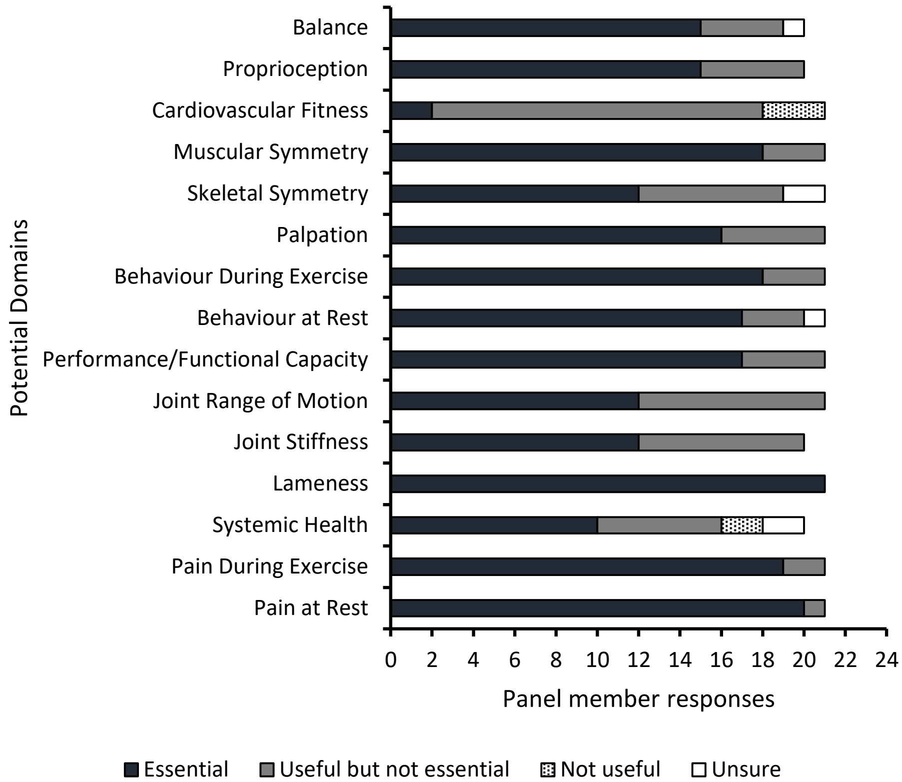

Figure 2: Expert opinion on domains to be included in an equine musculoskeletal outcome score.

3.3 Delphi step 3

The critical number required for the proportion in agreement (considering the domain to be essential) for a panel of 21 members according to Ayre and Scally [40] is 15 (71.4%), with a minimum CVR critical value of 0.429 [40]. Therefore using content validity ratios the number of possible domains for inclusion in TEMROS was reduced from 16 to 10. These were, with CVR values provided in parentheses: lameness (1.00), pain at rest (0.91), pain during exercise (0.81), behaviour during exercise (0.71), muscular symmetry (0.71), performance/functional capacity (0.62), behaviour at rest (0.62), palpation (0.52), balance (0.50) and proprioception (0.50). The domains with CVR critical values less than the required critical value were: joint stiffness (0.20), joint range of movement (0.14), skeletal symmetry (0.14), systemic health (0.00) and cardiovascular fitness (-0.81).

3.4 Delphi step 4

Seven panel members responded to the list of 10 domains positively and there were no further domains proposed for inclusion. There were three comments that centred on domains that should not be included. Three experts suggested that systemic health does not need to be measured within an outcome score, as this should be a pre-requisite for undertaking a rehabilitation programme and two mentioned cardio-vascular fitness measurement being outside the scope of a musculoskeletal assessment tool.

4. Discussion

Using expert's experience and opinion, this study aimed to develop a consensus on the domains to be included in a model for a composite outcome score for horses undergoing rehabilitation. These data indicate that observational data (e.g. lameness and behaviour due to pain) and hands-on (e.g. palpation on soft tissue) were considered essential for inclusion within a musculoskeletal OM. The broad range of domains in this study's model suggests that an outcome score needs to contain a variety of data. Indeed, this approach would provide a holistic view of the status of the horse undergoing therapy (figure 3).

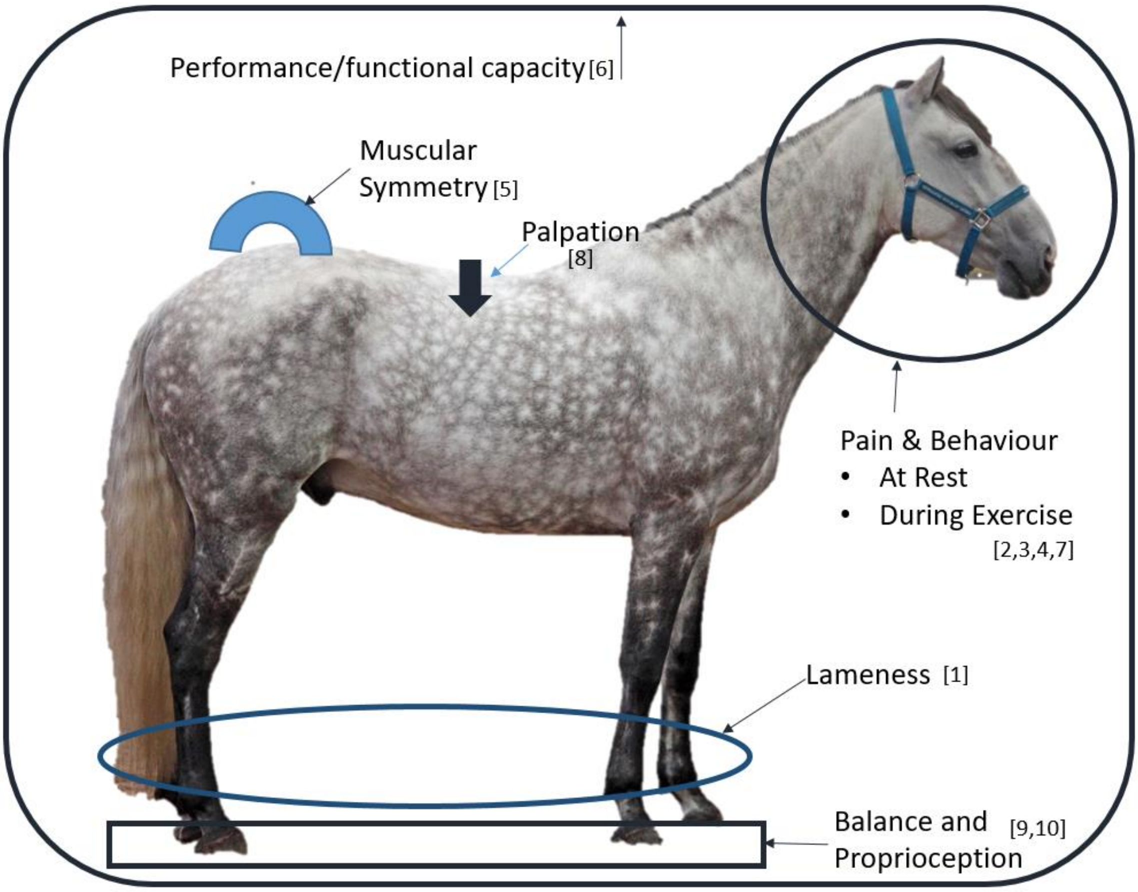

Figure 3: Ten domains for measurement, as agreed by the expert panel, to be included in outcome score for equine musculoskeletal rehabilitation. In order of highest agreement the domains (with number in square brackets) are: 1: lameness, 2: pain at rest, 3: pain during exercise, 4: behaviour during exercise, 5: muscular symmetry, 6: performance/functional capacity, 7: behaviour at rest, 8: palpation, 9: balance and 10: proprioception.

4.1 Lameness

The highest agreement across the panel was for the inclusion of a lameness measurement within TEMROS. In equine practice lameness is typically evaluated by observing movement asymmetry in trot, however this often presents a challenge especially in horses presenting with low grade lameness [41,42]. For gold standard detection and evaluation, force plates are recommended although these are not used outside the research environment and not practical for clinical assessment, therefore inertial sensor systems are useful where force plate analysis is not practical [43]. In practice, lameness assessment is commonly conducted by a visual gait assessment without technological equipment [41]. Visual assessment, without technological equipment, has been investigated for both intra- and inter-rater reliability. Keegan et al [42] studied the reliability of overground evaluation of lameness to determine if clinicians could agree on whether horses were lame and if so, which was the limb and score for the maximum level of lameness [42]. The American Association of Equine Practitioner (AAEP) scoring method was used, which is a 6 point scale where 0: Lameness not perceptible under any circumstances; 1: Lameness is difficult to observe and is not consistently apparent, regardless of circumstances (e.g. under saddle, circling, inclines, hard surface, etc.); 2: Lameness is difficult to observe at a walk or when trotting in a straight line but consistently apparent under certain circumstances (e.g. weight-carrying, circling, inclines, hard surface, etc.); 3: Lameness is consistently observable at a trot under all circumstances; 4: Lameness is obvious at a walk and 5: Lameness produces minimal weight bearing in motion and/or at rest or a complete inability to move (https://aaep.org/horsehealth/lameness-exams-evaluating-lame-horse). Keegan and colleagues [42] found that agreement of grading mild lameness was low (61.9%), although the agreement of lameness being present in horses scored greater than 1.5 on the AAEP scale was higher (93.1%) [42]. In addition, previous studies have shown lower agreement when practitioners assessed videos of lame horses [44,45,46]. Therefore it is suggested that multiple evaluators should not be used to evaluate lameness. In contrast to the AAEP score, one prominent equine veterinarian reported that too many horses with different levels of lameness have to be graded 3 on the AAEP scale and therefore in practice they use their own scale [47,48]. This recommended scale has 9 categories, where 0 = sound; 2 = mild; 4 = moderate; 6 = severe; 8 = non-weight bearing. The marked difference is that the grading system is applied in individual gaits and tests, for instance in a straight line or on a circle, to give a more accurate picture of the lameness, as it is their consensus that 0-5 represents insufficient grades and other systems using scores 0-10 consistent of too many options to be useable [47]. Whilst lameness was the domain which achieved universal agreement (100%), hence it should be included, how lameness evaluation is integrated remains challenging especially in the presence of bilateral lameness, lameness occurring only with specific conditions such as under saddle or in the case of an asymmetric gait that is due to morphology or laterality. The premise of an outcome score for practitioners is that it should be easy to use in clinical practice, therefore although technology may be increasingly available [49] whilst it is not yet in every practice or available to non-veterinary practitioners, a categorical subjective score would need to be included in TEMROS. The exact choice of grading system requires further study due to the absence of a universally accepted method that is easy to define, repeatable and can take into account the range of clinical presentations of lameness [47]. Until this is available physiotherapists should evaluate lameness individually, based on intra-rater reliability of lameness assessment being more reliable that inter-rater and that agreement between 'improvement' or 'worsening' in horses seen on multiple occasions is repeatable to use as an indicator of improvement, irrespective of the absolute score [46].

4.2 Pain assessment

Four domains selected related to the assessment of pain: pain at rest; behaviour at rest; pain during exercise and behaviour during exercise. Whilst crucial to horse welfare, the recognition and measurement of pain in horses is widely acknowledged to be difficult [50,51] due to pain levels reported by an observer being subjective and open to bias [52]. Pain has been reported to change facial expression in mice [53], rats [54] and more recently in horses via the horse grimace scale [50] and the equine pain face [51]. Both these equine scales have been validated for recording pain at rest by categorical scoring of facial expression and thus either could be used for the pain and behaviour at rest domains within TEMROS. The use of pain assessment for chronic, longer term pain conditions would have to be considered in the context of rehabilitation as this process takes longer than the duration of pain evaluation in the trials. These scoring systems have been shown to have acceptable inter-rater reliability for horses with acute pain. It would be of interest to know if veterinary professionals score similarly to the non-trained carers of horses undergoing treatment. Whether carers can objectively evaluate pain and not be altered by bias in either direction has not been reported nevertheless it is important to ensure that accurate pain assessment leads to optimal pain management throughout the whole course of treatment

Pain and behaviour during exercise could theoretically be integrated within TEMROS via scoring of facial expressions [55] and whole-horse behaviours during in hand and groundwork, and ridden work [56]. The level of activity that the horse was undertaking at the stage of rehabilitation would have to be factored into the outcome score, as early phase programmes may prohibit ridden activity, so pain and behaviour during handling tasks such as leading or ground work would need to be considered. As well as the task and the environment the assessment occurs in, an additional element that may alter horse's behaviour is the effect of the handler [57]. Therefore validity of pain assessment via facial expressions or whole horse behaviours during in-hand and groundwork with a handler, and in different locations such as an indoor arena or an outside location needs to be studied further.

It is of significant importance to horse welfare that the signs of pain in horses, whether in the stable or whilst being handled/ridden are considered during assessment. Evaluation of rehabilitation progress would not be holistic without including monitoring of pain, therefore further studies are required to test the application of pain assessment methods (e.g. Equine Pain face [51] or the ethogram for the assessment of pain in ridden horses [56]), specifically to rehabilitation programmes.

4.3 Muscle symmetry

The need to evaluate muscle symmetry is apparent when considering pathologies such as those in the region of the sacro-iliac joint, which may result in asymmetric atrophy of the overlying gluteus medius muscle [58]. Thoracolumbosacral pain can result in thoracic epaxial muscle wastage [7,59] which anecdotally may be lateralised and therefore asymmetric. Epaxial muscle size can be measured with ultrasound imaging [8,60,61] but this method may not always be accessible due to cost and its setting in veterinary or research laboratories. External muscle profile shape can be recorded with a low cost piece of equipment called a flexicurve ruler and this has been shown to be repeatable in the thoracic region [31] however the use of a flexicurve has not been reported on in other areas of the muscular system. The repeatability of a muscle scoring system devised by the authors of a study to investigate the relationship between thoracolumbar kinematics and muscle tone and tension in dressage horses found moderate agreement between five assessors (0.60-0.79) [62]. It was suggested that the muscle score could be used by physiotherapists to identify and monitor muscle development, however the authors' note the scale was subjective and only applicable to dressage horses. Therefore if this domain is to be included within TEMROS objective measures need to be further developed for clinical practice and tested for reliability and validity for horses in all equestrian disciplines, to be applicable to the possible range of horses undergoing rehabilitation.

4.4 Performance/functional capacity

Most tests of performance in horses have a strong physiological basis, such as standard exercise tests which evaluate relative speed and heart rate or blood lactate levels [63,64]. The intensity of the exercise effort in standard exercise tests, albeit submaximal, may not be appropriate for horses undergoing rehabilitation. A test of performance and functional capacity would need to be at lower exercise intensities and personalised to the stage of rehabilitation [65,66]. In human sports medicine function performance tests are used to evaluate return to play status in footballers [67], muscle strength and functional performance in recreational athletes following anterior cruciate ligament reconstruction [68] as well as function in patients with patella tendinosis or achilles tendinopathy [20,21]. Similarly, in dogs, functional tests are available such as the Canine Brief Pain Index and the Helsinki Chronic Pain Index [23,54] which include questions on tasks such as how well the dog rises to standing and willingness to walk or run. A functional score for dogs with neurological conditions has been tested for inter-rater reliability by seven observers scoring tasks of progressive difficulty such as standing up from lying, walking in turns or walking stairs [25]. The performance was graded with a numeric score from 0, indicating the dog cannot perform the task to 4, which represented normal motor function. No such scores exist in equine assessment but a simple battery of tests could be devised that included movements such as flexion of the neck [69] and turning small circles [70]. Any such testing procedure would need to be subject to evaluation of face and content validity and reliability testing similar to the neurological function tests for dogs devised by Boström et al [25].

4.5 Palpation

The panel agreed that palpation should be included in the proposed composite outcome score and it was expected that manual palpation would be required as local assessment of soft tissues and joint margins is commonly undertaken when assessing injury and pain [70]. Response to the manual palpation can be evaluated in the form of the behavioural response and/or evaluation of localised short-term change in the tissue being palpated, with a lower threshold to the onset of these responses indicative of a higher level of pain arising from these soft tissues [26-28,71,72]. Pain sensitivity, as a subjective experience, is individually variable in humans and based on complex physical and psychological interactions [73]; similarly third-party assessment of pain in animals has found wide intra-species variation exists as well as reported differences between species [74]. In horses, subjective judgement of pain thresholds by manual palpation is commonplace [28], therefore the use of quantitative tools to assess responses to palpation may be preferable to subjective pain assessment because this allows rating of response with a force output. Pressure algometry (PA) uses a calibrated pressure gauge to objectively record the threshold the onset of pain in the tissues it is applied on [77] The PA has been used to evaluate chiropractic interventions for equine thoracolumbar pain [11] and algometry measurements correlate with palpation scores (r = -0.90) [28]. However, reports that repeated PA application can result in sensitivity or habituation to the PA tool [29,71] could limit their validity in clinical practice. As an alternative, categorical scoring systems can be used to score response to manual palpation and use of this form of reporting could be integrated into TEMROS [9,27-29]. Merrifield-Jones et al [29] used a six-point score, where 0 is described as soft, low tone; 1 as normal; 2 as increased muscle tone but painful; 3 as increased muscle tone and/or painful (slight associated spasm on palpation, no associated movement; 4 painful (associated spasm on palpation with associated local movement, i.e. pelvis tilt, extension response) and 5 as very painful (spasm plus behavioural response to palpation, i.e. ears flat back, kicking). This score has shown excellent inter-rater reliability on a small sample of ten riding school horses between three physiotherapists when assessing epaxial soft tissue (ICC 0.09) [29]. The use of the PA tool, if practitioners were trained, could provide objective data if habituation and sensitisation were considered but the use of a categorical scale would provide a cost effective and convenient method of assessing response to palpation.

4.6 Balance and Proprioception

The final two domains that reached the minimal critical value for inclusion were balance and proprioception. The first study to measure balance in horses investigated postural sway using force platforms demonstrated that the standing horse has small movements of the centre of pressure resulting from small adjustments of muscle tension, indicating the stability of the quiet standing horse's centre of mass [75]. Whilst balance has not been measured in relation to musculoskeletal injury, motion of the centre of pressure does increase with medical sedation administered intravenously [76]. Signs of ataxia such as trembling, locking and unlocking of joints, weight shifts and obvious swaying were observed and it could be theorized that injury to one component required to maintain balance, such as sensory input, motor responses and cognitive processes [75] could have similar effects. To further examine potential clinical signs from neurological deficits, in relation to balance, twenty horses were blindfolded whilst stood on a force platform [77]. In these horses, movement amount and velocity increased, and showed greater within-trial variability when horses were blindfolded compared to their sighted measurements. Force platforms have been used as a primary outcome variable to assess the effects of osteoarthritis, surgically induced into the carpal joint in a group of 16 young horses [78]. Half of the cohort underwent an exercise regime on a water treadmill from 15 days following the surgery, five days a week for a total on ten weeks. At reassessment the horses that had been exercised on the water treadmill had significantly improved static balance control compared to control group of horses with carpal joint osteoarthritis. It should be noted that whilst these three force plate studies assessed postural sway during stance, gait involves spinal reflexes that might respond differently to effect balance during locomotion [76] therefore the results are limited as they cannot directly be translated to balance during gait.

Proprioception, as a domain listed to be included in TEMROS, does not have any objective measurement techniques reported for horses. However, postural stability relies on motor components of the musculoskeletal system to maintain balance and this includes proprioceptive information. Muscles induce joint motion and are also responsible to stabilising joints during motion therefore proprioceptive feedback is crucial to balance control [75,78}. Impairment to sensory and motor components, possibly due to joint injury, could affect postural control and if measured could also provide a proxy for proprioceptive deficit, but understanding this relationship within the scope of equine rehabilitation requires further analysis.

Force platforms could be used to measure balance and proprioceptive changes as a result of therapeutic interventions, although laboratory-based equipment is required because equine force platforms are not easily mobile. For clinical practice other methods to measure balance are necessary. Exercises to challenge balance and activate the trunk core muscles have been suggested as part of rehabilitation plans [79]. These exercises destabilise the horse by lifting a limb and inducing a weight transference to the contra- or ipsi-lateral weight-bearing limbs, however they do not have any form measurement to evaluate their effectiveness. A pressure mat that measures percentage weight distribution between limbs is available for canine orthopaedic assessment [80] and if a similar measurement method or a score system could be developed for horses then these positions could be used as a form of balance evaluation.

4.7 Limitations to the study

The number of experts selected to participate was small and was carried out based on the criteria (knowledge of research published and industry expertise) of the authors. This could present bias to the panel however once formed, TEMROS could be presented to the wider equine community for consideration and content validation. It would have been of benefit to have an understanding of the rationale for inclusion [81] to allow retrospective analysis of domains chosen. The high levels of agreement for the domains selected supports the consensus is based on common experience and practice.

Although a wide literature search has been completed to map potential reliable and valid measurement tools/tests to each domain it is possible that there are suitable tests/tools which were not suggested for inclusion by the panel. An example is thermography which has been used to measure surface temperature of racehorses' epaxial muscles in response to training [82]. Skin temperature measurements have not been used to evaluate effects of rehabilitation intervention however the reducing cost of thermography cameras may allow more horses to be imaged with this non-invasive and non-ionizing modality, albeit following strict protocols for carrying out and analysing results [83]. It should be noted that the choices of tests are evaluated in relation to those considered practical and feasible to use *ex vivo*. To be valid as a measure of rehabilitation outcome, each domain should have face validity which is a key factor in the development of an efficient OM is for the score in the absence of any gold standard [13].

4.8 TEMROS - further development

A composite score integrating the above domains takes into account several behaviours and physiological parameters by including scores for each specific parameter. There are domains that have various scoring systems or measurement tools, such as lameness and palpation and the final system/tool which require further testing to be validated. There are also domains where measurement techniques have yet to be designed for or tested, for instance muscle symmetry and proprioception, and therefore these areas need further development. Some of the parameters could be weighted according to perceived significance or they could be graded equally [33] and evaluation of this requires further development. However, TEMROS has the potential to provide a holistic assessment which would be relevant to rehabilitation of injury, as the whole horse is undergoing the rehabilitation not just the condition.

5. Conclusions

The Delphi methodology was successfully applied to attain consensus across the selected international expert panel that there is a need for an outcome measure for equine rehabilitation and agreement on the domains that such a measure should include. The expert panel agreed that lameness, pain at rest, pain during exercise, behaviour during exercise, muscular symmetry, performance/functional capacity, behaviour at rest, palpation, balance and proprioception should be included. The challenge going forward is to combine measures for each of these domains that are reliable, valid and easy to use in clinical practice. With reliably measured domains, and subsequent validity testing, TEMROS could provide a composite score that has equine practitioner consensus that could support clinical practice as well as substantiate treatment choices to improve horse welfare.

Author Contributions: Conceptualization, G.T. and J.W; methodology, G.T., K.N and J.W; writing—original draft preparation, G.T, K.N., J.F. and J.W.; writing—review and editing, G.T, K.N., J.F. and J.W.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding

Acknowledgments: The authors acknowledge the contribution of the experts to the Delphi processes

Conflicts of Interest: The authors declare no conflict of interest

References

1. Kaneps AJ. Practical rehabilitation and physical therapy for the general equine practitioner. *Veterinary Clinics: Equine Practice*. **2016** *Apr 1;32(1)*:167-80.
2. Tabor, G. Routine equine physiotherapy. *Equine Veterinary Education*. **2018** https://doi.org/10.1111/eve.12940
3. Haussler, K.K. The role of manual therapies in equine pain management. *Veterinary Clinics: Equine Practice*, **2010** *26*(3): 579-601.
4. Clayton, H.M. Core training and rehabilitation in horses. *Veterinary Clinics: Equine Practice*, **2016** *32*(1): 49-71.
5. Paulekas, R. and Haussler, K.K. Principles and practice of therapeutic exercise for horses. *Journal of Equine Veterinary Science*, **2009** *29*(12): 870-893.
6. Tabor, G. and Williams, J. The use of outcome measures in equine rehabilitation. *The Veterinary Nurse*, **2018** *9(9)*: 497-500.
7. Coomer, R.P., McKane, S.A., Smith, N. and Vandeweerd, J.M.E. A controlled study evaluating a novel surgical treatment for kissing spines in standing sedated horses. *Veterinary Surgery*, **2012** *41*(7): 890-897.
8. Stubbs, N.C., Kaiser, L.J., Hauptman, J. and Clayton, H.M. Dynamic mobilisation exercises increase cross sectional area of musculus multifidus. *Equine Veterinary Journal*, **2011** *43*(5): 22-529.
9. Schils, S.J. and Turner, T.A. Functional Electrical Stimulation for equine epaxial muscle spasms: retrospective study of 241 clinical cases. *Comparative Exercise Physiology*, **2014** *10*(2): 89-97.
10. Haussler, K.K., Hill, A.E., Puttlitz, C.M. and McIlwraith, C.W. Effects of vertebral mobilization and manipulation on kinematics of the thoracolumbar region. *American Journal of Veterinary Research*, **2007** *68*(5): 508-516.
11. Sullivan KA, Hill AE, Haussler KK. The effects of chiropractic, massage and phenylbutazone on spinal mechanical nociceptive thresholds in horses without clinical signs. *Equine Veterinary Journal*. **2008** *Jan;40(1)*: 14-20.
12. Outcome and Experience measures. Available online: https://www.csp.org.uk/professional-clinical/research-and-evaluation/outcome-and-experience-measures (accessed 16 Sept 2019)
13. Boers, M., Kirwan, J.R., Wells, G., Beaton, D., Gossec, L., d'Agostino, M.A., Conaghan, P.G., Bingham III, C.O., Brooks, P., Landewé, R. and March, L. Developing core outcome measurement sets for clinical trials: OMERACT filter 2.0. *Journal of Clinical Epidemiology*, **2014** *67*(7): 745-753.
14. Chiarotto, A., Terwee, C.B., Deyo, R.A., Boers, M., Lin, C.W.C., Buchbinder, R., Corbin, T.P., Costa, L.O., Foster, N.E., Grotle, M. and Koes, B.W., A core outcome set for clinical trials on non-specific low back pain: study protocol for the development of a core domain set. *Trials*, **2014** *15*(1): 511.
15. Jacklin, B.D., Minshall, G.J. and Wright, I.M. A new technique for subtotal (cranial wedge) ostectomy in the treatment of impinging/overriding spinous processes: description of technique and outcome of 25 cases. *Equine Veterinary Journal*, **2014** *46*(3): 339-344.
16. Walmsley, J.P., Pettersson, H., Winberg, F. and McEvoy, F. Impingement of the dorsal spinous processes in two hundred and fifteen horses: case selection, surgical technique and results. *Equine Veterinary Journal* **2014** *34*(1): 23-28.
17. Nagy A, Quiney L, Dyson S. Long-term outcome of 84 horses with sacroiliac joint region pain with (n= 69) or without (n= 15) other orthopaedic problems. Equine Veterinary Education. **2019.** Chiarotto, A., Deyo,

R.A., Terwee, C.B., Boers, M., Buchbinder, R., Corbin, T.P., Costa, L.O., Foster, N.E., Grotle, M., Koes, B.W. and Kovacs, F.M. Core outcome domains for clinical trials in non-specific low back pain. *European Spine Journal*, **2015** *24*(6): 1127-1142

18. Kyte, D.G., Calvert, M., Van der Wees, P.J., Ten Hove, R., Tolan, S. and Hill, J.C. An introduction to patient-reported outcome measures (PROMs) in physiotherapy. *Physiotherapy*, **2015** *101*(2): 119-125.
19. Visentini, P.J., Khan, K.M., Cook, J.L., Kiss, Z.S., Harcourt, P.R., Wark, J.D. and Victorian Institute of Sport Tendon Study Group. The VISA score: an index of severity of symptoms in patients with jumper's knee (patellar tendinosis). *Journal of Science and Medicine in Sport*, **1998** *1*(1): 22-28.
20. Robinson, J.M., Cook, J.L., Purdam, C., Visentini, P.J., Ross, J., Maffulli, N., Taunton, J.E. and Khan, K.M. The VISA-A questionnaire: a valid and reliable index of the clinical severity of Achilles tendinopathy. *British Journal of Sports Medicine*, **2001** *35*(5): 335-341.
21. Thorborg, K., Hölmich, P., Christensen, R., Petersen, J. and Roos, E.M. The Copenhagen Hip and Groin Outcome Score (HAGOS): development and validation according to the COSMIN checklist. *British Journal of Sports Medicine*, **2011** *45*(6): 478-491.
22. Hielm-Björkman, A.K., Kuusela, E., Liman, A., Markkola, A., Saarto, E., Huttunen, P., Leppäluoto, J., Tulamo, R.M. and Raekallio, M., Evaluation of methods for assessment of pain associated with chronic osteoarthritis in dogs. *Journal of the American Veterinary Medical Association*, **2003** *222*(11): 1552-1558.
23. Brown, D.C., Boston, R.C., Coyne, J.C. and Farrar, J.T. Development and psychometric testing of an instrument designed to measure chronic pain in dogs with osteoarthritis. *American Journal of Veterinary Research*, **2007** *68(6):* 631-637.
24. Boström, A.F., Hyytiäinen, H.K., Koho, P., Cizinauskas, S. and Hielm-Björkman, A.K. Development of the Finnish neurological function testing battery for dogs and its intra-and inter-rater reliability. *Acta Veterinaria Scandinavica*, **2018** *60(1):* 56.
25. Haussler KK, Erb HN. Pressure algometry for the detection of induced back pain in horses: a preliminary study. *Equine Veterinary Journal.* **2006** *38(1)*: 76-81
26. De Heus P, Van Oossanen G, Van Dierendonck MC, Back W. A pressure algometer is a useful tool to objectively monitor the effect of diagnostic palpation by a physiotherapist in warmblood horses. Journal of Equine Veterinary Science. **2010** 30(6): 310-321.
27. Varcoe-Cocks K, Sagar KN, Jeffcott LB, McGowan CM. Pressure algometry to quantify muscle pain in racehorses with suspected sacroiliac dysfunction. *Equine Veterinary Journal*. **2006** *38(6)*: 558-562
28. Merrifield-Jones, M., Tabor, G. and Williams, J. Inter and Intra-Rater Reliability of Soft Tissue Palpation in the Equine Thoracic Epaxial Region *Journal of Equine Veterinary Science* **2019** https://doi.org/10.1016/j.jevs.2019.102812
29. Tabor, G., Elliott, A., Mann, N. and Williams, J. Equine Posture Analysis: Development of a Simple Tool to Record Equine Thoracolumbar Posture. *Journal of Equine Veterinary Science*, **2019** *73:* 81-83
30. Greve, L. and Dyson, S. A longitudinal study of back dimension changes over 1 year in sports horses. *The Veterinary Journal.* **2015** Jan 1;203(1): 65-73.
31. Liljebrink, Y. and Bergh, A. Goniometry: is it a reliable tool to monitor passive joint range of motion in horses? *Equine Veterinary Journal*.**2010** *Nov;42*: 676-82.
32. Adair, H., Marcellin-Little, D. and, Levine, D. Validity and repeatability of goniometry in normal horses. *Veterinary and Comparative Orthopaedics and Traumatology* **2016** *29(04)*: 314-9
33. Gleerup KB, Lindegaard C. Recognition and quantification of pain in horses: A tutorial review. Equine Veterinary Education. **2016** 28(1): 47-57.
34. Klassen AF, Cano SJ, Schwitzer JA, Scott AM, Pusic AL. FACE-Q scales for health-related quality of life, early life impact, satisfaction with outcomes, and decision to have treatment: development and validation. *Plastic and Reconstructive Surgery*. **2015** *Feb 1;135(2*): 375-86.
35. Tabor, G. and Williams, J. Objective Measurement in Equine Physiotherapy. Comparative Exercise Physiology **2018** In Press
36. Hsu, C.C. and Sandford, B.A. The Delphi technique: making sense of consensus. *Practical Assessment, Research & Evaluation [Internet]. Available from: http://pareonline. net/getvn. 2007 asp*, *12:* 10.
37. Murray IR, Geeslin AG, Goudie EB, Petrigliano FA, LaPrade RF. Minimum information for studies evaluating biologics in orthopaedics (MIBO): platelet-rich plasma and mesenchymal stem cells. *JBJS*. **2017** 99(10): 809-19.
38. Lawshe, C. H. A quantitative approach to content validity. *Personnel psychology*, **1974** *28:* 563–575

39. Ayre, C. and Scally, A.J. Critical values for Lawshe's content validity ratio: revisiting the original methods of calculation. *Measurement and Evaluation in Counseling and Development*, **2014** *47*(1): 79-86.
40. Starke, S.D. and Oosterlinck, M. Reliability of equine visual lameness classification as a function of expertise, lameness severity and rater confidence. *Veterinary Record*, **2018** *Oct 9*.
41. 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. and Frees, K.E. Repeatability of subjective evaluation of lameness in horses. *Equine Veterinary Journal*, **2010** *42*(2): 92-97.
42. McCracken MJ, Kramer J, Keegan KG, Lopes M, Wilson DA, Reed SK, LaCarrubba A, Rasch M. Comparison of an inertial sensor system of lameness quantification with subjective lameness evaluation. *Equine Veterinary Journal*. **2012** *Nov;44(6)*: 652-6.
43. Keegan, K.G., Wilson, D.A., Wilson, D.J., Smith, B., Gaughan, E.M., Pleasant, R.S., Lillich, J.D., Kramer, J., Howard, R.D., Bacon-Miller, C. and Davis, E.G. Evaluation of mild lameness in horses trotting on a treadmill by clinicians and interns or residents and correlation of their assessments with kinematic gait analysis. *American Journal of Veterinary Research*, **1998** 59(11): 1370-1377.
44. 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): 852-858.
45. Fuller, C.J., Bladon, B.M., Driver, A.J. and Barr, A.R. The intra-and inter-assessor reliability of measurement of functional outcome by lameness scoring in horses. *The Veterinary Journal*, **2006** *171(2):* 281-286.
46. Dyson, S. Can lameness be graded reliably? *Equine Veterinary Journal*, **2011** *43(4):* 379-382.
47. 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:* 64-70.
48. Bragança, F.S., Rhodin, M. and van Weeren, P.R. On the brink of daily clinical application of objective gait analysis: What evidence do we have so far from studies using an induced lameness model? *The Veterinary Journal*, **2018** *234:* 11-23.
49. 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.
50. Gleerup, K.B., Forkman, B., Lindegaard, C. and Andersen, P.H. An equine pain face. *Veterinary Anaesthesia and Analgesia*, **2015** *42*(1): 103-114.
51. Bussieres, 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): 294-306.
52. Langford DJ, Bailey AL, Chanda ML, Clarke SE, Drummond TE, Echols S, Glick S, Ingrao J, Klassen-Ross T, LaCroix-Fralish ML, Matsumiya L. Coding of facial expressions of pain in the laboratory mouse. *Nature methods*. **2010** *Jun;7(6)*: 447.
53. Sotocina SG, Sorge RE, Zaloum A, Tuttle AH, Martin LJ, Wieskopf JS, Mapplebeck JC, Wei P, Zhan S, Zhang S, McDougall JJ. The Rat Grimace Scale: a partially automated method for quantifying pain in the laboratory rat via facial expressions. *Molecular Pain*. **2011** *Jul 29;7:* 1744-8069.
54. 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:* 7-12.
55. Dyson S, Berger J, Ellis AD, 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*: 47-57.
56. Keeling LJ, Jonare L, Lanneborn L. Investigating horse–human interactions: The effect of a nervous human. *The Veterinary Journal*. **2009***;181(1)* :70-1.
57. Barstow, A. and Dyson, S. Clinical features and diagnosis of sacroiliac joint region pain in 296 horses: 2004–2014. *Equine Veterinary Education*, **2015** *27*(12): 637-647.
58. Zimmerman, M., Dyson, S. and Murray, R. Close, impinging and overriding spinous processes in the thoracolumbar spine: The relationship between radiological and scintigraphic findings and clinical signs. *Equine Veterinary Journal*, **2012** *44*(2): 178-184.
59. de Oliveira, K., Soutello, R.V., da Fonseca, R., Costa, C., Paulo, R.D.L., Fachiolli, D.F. and Clayton, H.M. Gymnastic training and dynamic mobilization exercises improve stride quality and increase epaxial muscle size in therapy horses. *Journal of Equine Veterinary Science*, **2015** *35*(11-12): 888-893.

60. Abe, T., Kearns, C.F. and Rogers, B. Reliability of ultrasound-measured muscle thickness of the longissimus dorsi muscle in thoroughbreds. *Comparative Exercise Physiology*, **2012** *8*(3-4): 189-194.
61. Walker VA, Tranquille CA, Dyson SJ, Spear J, Murray RC. Association of a subjective muscle score with increased angles of flexion during sitting trot in dressage horses. *Journal of Equine Veterinary Science. 40*: 6-15.
62. Sloet MV, Barneveld A. Comparison of the workload of Dutch warmblood horses ridden normally and on a treadmill. *The Veterinary Record*. **1995**; *137(6)*: 136-9.
63. Couroucé A. Field exercise testing for assessing fitness in French Standardbred trotters. *The Veterinary Journal*. **1999** *1;157(2)*: 112-22.
64. Munk, R., Møller, S. and Lindner, A. Effects of training with different interval exercises on horses used for show jumping. *Comparative Exercise Physiology*, **2013** *9*(1): 33-41.
65. Munsters, C.C., van Iwaarden, A., van Weeren, R. and van Oldruitenborgh-Oosterbaan, M.M.S. Exercise testing in Warmblood sport horses under field conditions. *The Veterinary Journal*, **2014** *202(1): 1*1-19.
66. Gomez-Piqueras, P., Gonzalez-Rubio, J., De Baranda, P.S. and Najera, A. Use of functional performance tests in sports: Evaluation proposal for football players in the rehabilitation phase. *Turkish Journal of Physical Medicine & Rehabilitation* **2018** *2587-0823*, *64*(2).
67. Larsen, J.B., Farup, J., Lind, M. and Dalgas, U. Muscle strength and functional performance is markedly impaired at the recommended time point for sport return after anterior cruciate ligament reconstruction in recreational athletes. *Human Movement Science*, **2015** *39:* 73-87.
68. Clayton HM, Kaiser LJ, Lavagnino M, Stubbs NC. Dynamic mobilisations in cervical flexion: Effects on intervertebral angulations. *Equine Veterinary Journal.* **2010** *42:* 688-94.
69. Goff, L. Physiotherapy assessment for animals. In: McGowan CM, Goff L, editors. *Animal Physiotherapy*. Second Edition, 2016 Wiley-Blackwell, West Sussex UK. Chapter 11, pp. 171–96
70. Haussler KK, Erb HN. Mechanical nociceptive thresholds in the axial skeleton of horses. *Equine Veterinary Journal*. **2006** *38(1)*: 70-75
71. Menke, E.S., Blom, G., van Loon, J.P. and Back, W. Pressure algometry in Icelandic horses: interexaminer and intraexaminer reliability. *Journal of Equine Veterinary Science*, **2016** *36:* 26-31.
72. Coghill, R.C. Individual differences in the subjective experience of pain: new insights into mechanisms and models. *Headache: The Journal of Head and Face Pain*, **2010** *50(9):* 1531-1535.
73. Stasiak, K.L., Maul, D.O.N., French, E., Hellyer, P.W. and Vandewoude, S. Species-specific assessment of pain in laboratory animals. *Journal of the American Association for Laboratory Animal Science*, **2003** *42(4):* 13-20.
74. Clayton, H.M., Bialski, D.E., Lanovaz, J.L. and Mullineaux, D.R. Assessment of the reliability of a technique to measure postural sway in horses. *American Journal of Veterinary Research*, **2003** *64*(11): 1354-1359.
75. Bialski, D., Lanovaz, J.L., Bohart, G.V., Mullineaux, D.R. and Clayton, H.M. Effect of detomidine on postural sway in horses. *Equine and Comparative Exercise Physiology*, **2004** *1(1):* 45-50.
76. Clayton, H.M. and Nauwelaerts, S. Effect of blindfolding on centre of pressure variables in healthy horses during quiet standing. *The Veterinary Journal*, **2014** *199*(3): 365-369.
77. King, M.R., Haussler, K.K., Kawcak, C.E., McIlwraith, C.W. and Reiser II, R.F. Effect of underwater treadmill exercise on postural sway in horses with experimentally induced carpal joint osteoarthritis. *American Journal of Veterinary Research*, **2013** *74(7):* 971-982.
78. Stubbs NC, Clayton HM. *Activate Your Horse's Core: Unmounted Exercises for Dynamic Mobility, Strength, & Balance.* Sport Horse Publications; 2008.
79. Clough, W.T. and Canapp Jr, S.O. Assessing Clinical Relevance of Weight Distribution as Measured on a Stance Analyzer through Comparison with Lameness Determined on a Pressure Sensitive Walkway and Clinical Diagnosis. *Veterinary and Comparative Orthopaedics and Traumatology*, **2018** *31*(S 02): A3622.
80. Basch E, Spertus J, Dudley RA, Wu A, Chuahan C, Cohen P, Smith ML, Black N, Crawford A, Christensen K, Blake K. Methods for developing patient-reported outcome-based performance measures (PRO-PMs). *Value in Health*. **2015** *18(4)*: 493-504.
81. Soroko M, Henklewski R, Filipowski H, Jodkowska E. The effectiveness of thermographic analysis in equine orthopedics. Journal of equine veterinary science. **2013** 33(9): 760-2.
82. Soroko M, Howell K. Infrared thermography: current applications in equine medicine. Journal of Equine Veterinary Science. **2018** 60: 90-6.