The Old School, Brewhouse Hill, Wheathampstead,
Hertfordshire AL4 8AN, UK
www.ufaw.org.uk

Animal Welfare 2014, 23: 323-341
ISSN 0962-7286
doi: 10.7120/09627286.23.3.323

Equine on-farm welfare assessment: a review of animal-based indicators

E Dalla Costa*, L Murray, F Dai, E Canali and M Minero

Università degli Studi di Milano, Dipartimento di Scienze Veterinarie e Sanità Pubblica, Via Celoria 10, 20133 Milano, Italy
* Contact for correspondence and requests for reprints: emanuela.dallacosta@unimi.it

Abstract

The adaptability of horses and donkeys to different types of activity has seen the European equine industry become an important economic sector, giving rise to increasing concern regarding equine welfare. As part of the AWIN (Animal Welfare Indicators) project, this review focuses on scientific literature to find potential animal-based welfare indicators — the initial step in developing a valid, reliable and feasible on-farm welfare assessment protocol for equines. Forty-nine indicators were considered and classified in accordance with the four Principles and twelve Criteria developed by Welfare Quality®. Only practical indicators specifically for on-farm use were included, those requiring the use of specific instruments or laboratory analysis were excluded. Academic scientists, partners and collaborators of the AWIN project, discussed and agreed on validity, reliability, on-farm feasibility and acceptance by farmers for each indicator. Some aspects of equine welfare have been thoroughly investigated and appear to have indicators ready for on-farm use (eg 'absence of prolonged hunger', 'absence of injuries and diseases'). On the other hand, a lack of animal-based measures were identified for other Criteria such as 'absence of pain' and 'positive emotional state'. Ongoing research within the AWIN project has begun exploring some of the aforementioned Criteria — these preliminary results of promising indicators have been included (eg Horse Grimace Scale and Qualitative Behaviour Assessment). Further research should address the validity and reliability of indicators, such as human-animal relationship tests and signs of cold stress. As well as for working equines, the development and application of a welfare assessment protocol could be the first step for enhancing on-farm equine welfare.

Keywords: *animal-based indicator, animal welfare, donkey welfare, equine welfare, horse welfare, on-farm welfare assessment protocol*

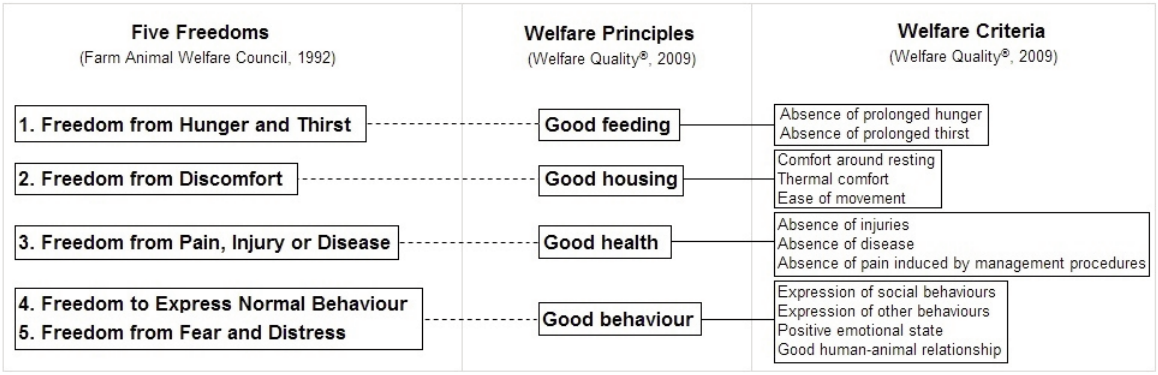
Introduction

It is estimated that more than six million equines live in Europe, however there are no definitive statistics (European Horse Network 2010; Faostat 2011). The European equine industry is an important economic sector, thanks to the adaptability of horses and donkeys to very different types of activity (eg breeding, leisure and sport, education) and the effect of people's continued fascination with equids, and their willingness to spend money on them as either a business or hobby. Equine welfare is an increasing cause for concern due to limitations of the present European legislation, which differs between countries and does not encompass all aspects of welfare. There is currently increased public awareness and demand for improved equine welfare (Fraser 2001). The frequent need for rapid responses to address contingent equine welfare issues and to answer public concerns has forced scientists to produce suboptimal criteria to assess welfare on-farm (Broom 2011).

Animal welfare "is a term that describes a potentially measurable quality of a living animal at a particular time and hence is a scientific concept" (Pond *et al* 2011). The assessment of animal welfare requires a multi-dimensional approach (Mason & Mendl 1993), and should aim to determine the actual welfare of animals, including both physical and mental states (EFSA Panel on Animal Health and Welfare 2012).

Funded by the European Commission, in the Seventh Framework Programme, the AWIN (Animal Welfare Indicators) project's goals include the improvement of animal welfare by developing practical on-farm welfare assessment protocols for several species, including horses and donkeys. This current review of scientific literature is the starting point to identify promising animal-based indicators. Based on the findings in this review, AWIN scientists will develop a research action plan to address the lack of knowledge regarding the validity, repeatability and feasibility of single indicators. The resulting list of indicators will then be tested on-farm by trained assessors. The overall assessment of welfare should be regarded as a multidimensional process that takes into consideration several aspects that are almost independent (eg good human-animal relationship and absence of pain). Due to the differences in equine use, housing and management throughout Europe, it should be clarified that the term 'on-farm' refers to any type of facility housing equines where the assessment may take place, where it is performed

Figure 1



The link between the concept of the Five Freedoms proposed by the Farm Animal Welfare Council and the four animal welfare Principles and 12 Criteria formulated by Welfare Quality®.

'on-site', such as riding schools, racecourses, stables, livery yards, sanctuaries, actual farms, etc. Working equines refers to animals used to transport crops, wood for fuel, water, building materials and people by carts or on their back, tillage and in occupational therapy programmes (Mekuria & Abebe 2010). It should also be clarified that 'farmer' refers also to the owner or primary carer of the animals.

In 2008, the EU Welfare Quality® project defined four welfare Principles, linked to twelve Criteria (Blokhuis *et al* 2010; Rushen *et al* 2011), starting from the concept of the animals' Five Freedoms (Brambell 1965) (see Figure 1). Each Principle is phrased to communicate a key welfare question and is divided into different Criteria. Each welfare Criterion represents a specific area of welfare, which indicates an area of concern; consequently, Criteria are independent of each other and form 'an exhaustive, but concise list' (Welfare Quality® Consortium 2009).

Two broad categories of indicator can be used to assess animal welfare at the farm level: animal-based and resource-based (EFSA Panel on Animal Health and Welfare 2012). The use of animal-based indicators for equine welfare assessment allows assessment of welfare in different housing conditions as the indicators refer to the animal itself rather than the environment. One important advantage of using animal-based indicators is the possibility of evaluating the animals, either by observation or by inspection. The advantages of using directly evaluable indicators are given by EFSA:

> animal-based measures are linked to welfare-related outcomes and they can be considered as a form of toolbox from which to select the range of measures necessary to address the specific objectives of the assessment for that particular species and category of animal at that time. That is to say, the measures chosen should be 'fit for purpose'. Which measure is the most appropriate will depend on a number of different things, e.g. the purpose of the assessment, the skills of the person collecting the measure, the conditions under which it is to be gathered, the time available to collect it and financial constraints" (EFSA Panel on Animal Health and Welfare 2012).

The research question addressed by this review was: to date, which animal-based indicators used to evaluate equine welfare are valid, reliable and feasible on-farm?

Materials and methods

To capture as many relevant citations as possible, a range of databases (Web of Science, CAB Abstracts, PubMed and Scopus) were searched to identify key studies addressing animal-based welfare indicators in equines since 1980 (see Table 1 for keywords used).

The search resulted in 4,940 citations from which relevant studies were selected for the review: we aimed to include key studies in equines that address animal-based welfare indicators in any housing condition and category (eg working equines). We included full papers published in peer-reviewed journals and proceedings and, when full papers were not available, abstracts written in English with a clear explanation of the experimental design and the methods followed were sought; we excluded any studies that solely concentrated on resource-based (eg bedding) or management-based (eg questionnaires) indicators. After exclusions, 54 papers from 21 countries remained, published between 1988 and 2013, which were relevant to the question posed for the review.

After studies had been selected, they were classified in tables according to the Five Freedoms (Brambell 1965; Farm Animal Welfare Council 1979), four Principles and twelve Criteria (Welfare Quality® Consortium 2009) (see Figure 1). Tables included information on animal category (age, sex, breed and attitude), type of housing (individually or group stabled, paddock, pasture, etc), sample size and validity, reliability and on-farm feasibility of the identified animal-based indicator, as well as references to the literature.

Validity concerns the relationship between an indicator and what it is supposed to measure or predict (Acock 2008). Criterion-related validity picks one or more criteria or standards for evaluating a scale, such as a predictive or a concurrent measure. Predictive validity measures the ability of an indicator to predict some later criterion, while concurrent validity measures the correlation between an indicator and other measures to which it is theoretically related (Kamphaus & Frick 2005). Reliability refers to repeatability in time and consistency within and between observers (Martin & Bateson 2007). On-farm feasibility considers the

https://doi.org/10.7120/09627286.23.3.323 Published online by Cambridge University Press

practical likelihood of using the indicator during on-farm inspection. Therefore, it is a more dynamic concept, dependent on factors such as the purpose of the assessment and budgetary constraints. Together with biosecurity and safety issues, time needed to collect the data as well as farmers' and stakeholders' acceptance, these comprise the main variables to be evaluated (Knierim & Winckler 2009).

Thirteen academic scientists, internationally acknowledged for their expertise in equine welfare and peer-reviewed publications on relevant topics, were selected as partners and collaborators in the AWIN project. They were given a fixed definition for validity, reliability, on-farm feasibility and were subsequently asked to fill in the tables, scoring each indicator within each paper on the above-mentioned variables. Possible scores were: tested/not tested (ie was the repeatability tested?); and yes/no (ie repeatable/not repeatable). Scientists agreed on definitions and scores of validity and reliability, whereas a consensus had to be reached regarding the on-farm feasibility of some indicators. The point of view of each scientist was considered, discussed and compared during a meeting; definitions and explanations were used to reach a consensus regarding on-farm feasibility and to define promising indicators to be included in the equine welfare assessment protocols. A research action plan was defined to cover the lack of knowledge for some of the indicators.

Table 1 Keywords used for database search.

Major descriptors	Combined with
Welfare	Equine
Welfare measure	Equid
Welfare indicator	Equus
Welfare assessment	Horse
Absence of prolonged hunger	Donkey
Absence of prolonged thirst	
Comfort around resting	
Thermal comfort	
Ease of movement	
Absence of injuries	
Absence of disease	
Absence of pain induced by management procedures	
Expression of social behaviours	
Expression of other behaviours	
Positive emotional state	
Good human-animal relationship	
Disease	
Pain	
Fear	
Discomfort	
Anxiety	
Frustration	
Stress	
Stress assessment	
Behaviour test	
Behaviour preference	
Preference test	
Body Condition Score	
Human-animal relationship	
Aggressive behaviour	
Aggression	
Learned helplessness	
Conflict behaviour	
Skin lesions	

Results and discussion

A total of 54 peer-reviewed papers regarding experimental studies on the development of animal-based welfare indicators satisfied the search criteria, identifying 49 indicators. The total number of recognised indicators seems large; however, following the evaluation of validity, reliability and feasibility, only a few meet all of the necessary requirements. As reported above, the discussion on equine animal-based indicators is presented following the four welfare Principles and twelve Criteria of Welfare Quality®.

Principle: good feeding

Animal-based indicators to assess the Principle good feeding, their validity, reliability and on-farm feasibility are reported in Table 2.

Criterion: absence of prolonged hunger

When dealing with horses and donkeys, particularly in Europe, obesity is as much a welfare issue as being of low weight; for example, some ponies or donkeys might be obese but still hungry. Therefore, we focused on animal-based indicators which would assess the appropriate nutrition of equines. Two categories of animal-based indicator were identified: i) weight estimation; ii) the feeling of hunger linked with behavioural expression by equines.

Weight estimation

The weight estimation of horses can be assessed by various methods: weigh tape, weight estimation formula, visual estimate and Body Condition Score (BCS). A weigh tape is a tool which is frequently used to record weight directly, by passing it around the horse at the lowest point of the withers. There are different commercially available weigh tapes with varying efficacy (Ellis & Hollands 1998, 2002). Weight estimation formulae use the heart girth and body length measurements in centimetres to calculate the weight in kilograms. There are a number of weight estimation formulae for horses and a separate one for donkeys (Carrol & Huntington 1988; Ellis & Hollands 1998; Burden 2012).

Table 2 Animal-based indicators for assessing the Principle good feeding.

Animal-based welfare indicators	Species	Housing/ category	Validity	Test-retest reliability	Inter-observer reliability	On-farm feasibility	References
Absence of prolonged hunger							
Weight estimation tape	H	S, P	No	–	–	–	Ellis & Hollands 1998, 2002
Weight estimation formula	H, D	S, P	Yes	–	–	–	Carrol & Huntington 1988; Ellis & Hollands 1998; Burden 2012
Visual estimate	H	S, P	No	–	–	–	Ellis & Hollands 1998; Reavell 1999; Burkholder 2000
Body Condition Score	H, D	S, P, W	Yes	Yes	Yes	Yes	Pearson & Ouassat 1996; Carrol & Huntington 1988; Burkholder 2000; Pritchard *et al* 2005; Burn *et al* 2009, 2010; Mekuria & Abebe 2010; Burden 2012; Cappai *et al* 2013; Quaresma *et al* 2013
Bedding investigation	H	S	–	–	–	–	Ninomiya & Kusunose 2004; Ninomiya *et al* 2007a
Bedding eating	H	S	–	–	–	–	Ninomiya & Kusunose 2004; Ninomiya *et al* 2007a
Resting behaviour	H	S	–	–	–	–	Ninomiya & Kusunose 2004; Ninomiya *et al* 2007a
Absence of prolonged thirst							
Skin tent test	H, D	W	No	Yes	No	Yes	Pritchard *et al* 2005, 2006, 2007, 2008; Burn *et al* 2009
Mucous membrane dryness	H, D	W	No	–	Yes	Yes	Pritchard *et al* 2005, 2008; Burn *et al* 2009; Mekuria & Abebe 2010
Drinking test	H, D	W	Yes	–	–	–	Pritchard *et al* 2006, 2008

H: Horse; D: Donkey; S: Single box; P: Paddock; W: Working equine.
Yes: tested and valid, reliable or feasible; No: tested and not valid, not reliable or not feasible; not tested.

Visual estimation appears to be the most commonly used method by experienced horse persons and veterinarians for determining equine weight (Ellis & Hollands 1998; Reavell 1999), it is a wholly subjective method using only visual appraisal. BCS is a well known and widely used method for assessing appropriate nutrition of farm animals, including equines (Huntington 1988; Pearson & Ouassat 1996; Pritchard *et al* 2005; Burn *et al* 2009, 2010; Carrol & Mekuria & Abebe 2010; Burden 2012). It is a subjective, semi-quantitative method for evaluating body fat and muscle that takes into account the deposition of body fat in different areas by separate examination of the neck, back, ribs, pelvis and rump (Carrol & Huntington 1988). Burden (2012) reported that:

> body condition scoring for donkeys and mules requires a different technique to that used in horses, as donkeys lay down fat stores in more localised areas and have a different body shape.

In horses, BCS assessment can be performed visually, through palpation, or both, while in donkeys, palpation is necessary due to the different length of the coat and thickness of the skin (Cappai *et al* 2013). It can be scored using a 5-point (Carrol & Huntington 1988; The Donkey Sanctuary BCS scale) or 9-point scale (Henneke *et al* 1983). The optimum body condition score is considered to be a 3 or a 5, with the 5- or the 9-point scale, respectively.

Weight estimation formulae were found to be valid for estimating weight (Carrol & Huntington 1988; Ellis & Hollands 1998; Burden 2012; Cappai *et al* 2013). However, weigh tapes were not. Compared with a weigh-bridge, estimates obtained using Spillers and Dalton weigh tapes were not accurate. Measures obtained were influenced by the dimensions of the horse, particularly whether it was greater or less than 15 hh (about 152 cm) (Ellis & Hollands 1998). Visual estimation of a horse's weight has been found to be inaccurate and unreliable (Ellis & Hollands 1998), particularly for the excessive subjectivity of the estimates (Burkholder 2000).

Burkholder (2000) reported that BCS is a repeatable measure when performed in accordance with specific protocols, and it also has good inter-observer reliability. Using the 5-point scale, BCS seemed to be reliable among 6–10 different observers (Pritchard *et al* 2005; Burn *et al* 2009, 2010). Test-retest reliability of the other indicators, as well as their repeatability, has never been evaluated.

On-farm feasibility has only been considered for BCS, which is feasible to measure with relative ease under different housing and management conditions: not only on-farm, but also in working equines (Carrol & Huntington

https://doi.org/10.7120/09627286.23.3.323 Published online by Cambridge University Press

1988; Pritchard *et al* 2005; Mekuria & Abebe 2010). Furthermore, the BCS system has been reported to be easy to learn and even the most complicated and evolved BCS protocol can be mastered relatively quickly through application and practice (Burkholder 2000). All the animal-based indicators found seem to be acceptable to farmers, because they require simple measurements.

In view of all the observations reported above, BCS is a valid, reliable, feasible and easy to learn, animal-based indicator; therefore it is probably the best current method of assessing on-farm nutrition of equines. Besides the quantity and quality of feed, one must take into account that other factors, as the age of the subject and the presence of disease can affect body condition.

The feeling of hunger

Food restriction — and the consequent eating frustration — might be necessary to improve the welfare of equines in the long term, in cases of specific clinical conditions or to treat obesity. Excellent body condition in a horse or a donkey does not necessarily mean that eating/grazing/foraging need is fulfilled. High energy diets provide the nutritional requirements, but a psychological need to forage for many hours per day may still exist. The feeling of hunger, as well as feeding satisfaction in the subject can be assessed using behavioural indicators such as bedding investigation and eating, and resting behaviour after a meal. Bedding investigation (smelling bedding or moving it with the nose) and bedding eating during the two hours post-feeding have been reported to be an indicator of eating frustration, linked with the feeling of hunger in horses (Ninomiya & Kusunose 2004), while resting behaviour after eating (eg standing-sleep) was described by Ninomiya *et al* (2007a,b) as an indicator of eating satisfaction in horses. However, the validity of bedding investigation, bedding eating and resting behaviour as behavioural indicators of eating frustration and satisfaction has never been studied and should be carefully evaluated. Additionally, there are doubts about the feasibility of using these indicators because they require a long observation time.

Further studies should be conducted to evaluate the validity and reliability of behavioural indicators (eg bedding eating) as these indicators could contribute important information regarding eating frustration (for stereotypies see the *Criterion: expression of other behaviours*) and feeding satisfaction that are considered to be welfare issues, primarily for stabled horses.

Criterion: absence of prolonged thirst

Indicators found to assess the absence of prolonged thirst were the skin tent test, mucous membrane dryness and the drinking test. All the animal-based indicators found have only been assessed in working equines, while in stabled equines, resource-based indicators are regularly preferred. Two categories of animal-based indicators were also identified: i) dehydration; ii) the feeling of thirst.

Dehydration

Dehydration can be assessed by performing the skin tent test or by checking for mucous membrane dryness. The skin tent test is assessed by pinching and immediately releasing the skin of the cranial margin of the animal's scapula and a vertical fold of skin overlying the *Musculus brachiocephalicus*, then observing when the skin returns to its normal position. If there is a delay in the return of tented skin to its normal position, the animal could be dehydrated (Pritchard *et al* 2005, 2006, 2008; Burn *et al* 2009). Mucous membrane dryness is evaluated using a fast filter paper placed on the gingival mucosa for 10 s (Pritchard *et al* 2008). Qualitative assessment of dryness and adhesion to the mucosa is scored with a 0–5 point scale.

The validity of the skin tent test has been evaluated in a number of studies (Pritchard *et al* 2005, 2006, 2008; Burn *et al* 2009), but was found to be limited, particularly for assessing dehydration in horses (there is a poor correlation with physiological measures such as plasma osmolarity). It is of paramount importance that the authors of these studies could not exclude the presence of the confounding effects of malnutrition. The skin tent test is a moderately repeatable measure (Pritchard *et al* 2007). Researchers found differences between different anatomical locations (eg side: skin tents on the left side of the animal were longer than on the right). Inter-observer reliability of the skin tent test can be improved with increased training of assessors (introducing the concept of biological variability, eg for elasticity of the skin) and by using a simplified score (yes–no) (Pritchard *et al* 2007). Although relatively simple and feasible, the qualitative and quantitative assessment of mucous membrane dryness does not seem to be a valid measure of dehydration. A study concerning inter-observer reliability of mucous membrane dryness considered it not to be reliable because, for example, drinking water can influence this measure by decreasing dryness (Pritchard *et al* 2006). A study on inter-observer reliability of mucous membrane dryness evaluates it as ambiguous in both horses and donkeys (Burn *et al* 2009).

The feeling of thirst

The drinking test is a simple experiment in which the assessor offers water-filled buckets (at ambient environmental temperature) to the animal, and observes its behaviour for 10 min (Pritchard *et al* 2006, 2008). To avoid bias due to other confounding factors, the bucket and the water provided should be familiar to the animal. It could be an easy way to assess the feeling of thirst in horses and donkeys, especially if they do not have free access to water.

The drinking test appears to be a valid, direct, animal-based measure for assessing the feeling of thirst, in particular, for horses exercising in conditions of high ambient temperature. Water intake also appears to be linked with dehydration of the subject (Pritchard *et al* 2006). However, possible confounding factors arising when testing exhausted horses, horses in a novel environment or when different motivation factors are present, should be noted and the results regarded with due caution.

Table 3 Animal-based indicators for assessing the Principle good housing.

Animal-based welfare indicators	Species	Housing/ category	Validity	Test-retest reliability	Inter-observer reliability	On-farm feasibility	References
Comfort around resting							
Lying behaviour	H	S	–	–	–	No	Heleski *et al* 2002; Pedersen & Ladewig 2004; Chaplin & Gretgrix 2010
Thermal comfort							
Behavioural signs of heat stress (increased frequency and depth of respiration, flared nostrils, profuse sweating, head nodding and apathy	H, D	W	Yes	Yes	Yes	Yes	Pritchard *et al* 2005, 2006; Minka & Ayo 2007; Burn *et al* 2009; Holcomb *et al* 2013
Behavioural signs of cold stress (shivering, shelter seeking, huddling)	H	P	Yes	–	–	Yes	Mejdell & Bøe 2005; Heleski & Murtazashvili 2010b
Ease of movement							
Daily activity	H	S	–	–	–	No	Chaplin & Gretgrix 2010
Locomotor stereotypies (box walking, weaving, fence pacing, pawing, box kicking)	H	S	–	–	–	Yes	McGreevy *et al* 1995; Heleski *et al* 2002; Bachmann *et al* 2003; Ninimiya *et al* 2007b

H: Horse; D: Donkey; S: Single box; P: Paddock; W: Working equine.
Yes: tested and valid, reliable or feasible; No: tested and not valid, not reliable or not feasible; not tested.

The reliability of the drinking test and repeatability of mucous membrane test have never been assessed. The drinking test appears to be a feasible animal-based indicator, however it is important to evaluate drinking test feasibility in relation to the condition of an on-farm welfare assessment protocol. Another major consideration regarding feasibility of the drinking test is the potential issue of biosecurity and the transfer of pathogens and disease among equines within facilities.

The difficulty of finding a valid and feasible animal-based measure for assessing absence of prolonged thirst is clear. On-farm feasibility and reliability for the drinking test should be investigated. At present, resource-based indicators, such as continuous water availability and cleanliness of drinkers, are the most valid, reliable and feasible indicators for on-farm assessment of this Criterion.

Principle: good housing

Animal-based indicators to assess the Principle good housing, their validity, reliability and on-farm feasibility are reported in Table 3.

Criterion: comfort around resting

The only animal-based indicator found to assess this Criterion was lying behaviour (Heleski *et al* 2002; Pedersen & Ladewig 2004; Chaplin & Gretgrix 2010). To achieve paradoxical sleep, horses prefer to lie down in lateral rather than in sternal recumbency (Pedersen & Ladewig 2004). For this reason, the inability to lie down affects their welfare and performance. Raabymagle and Ladewig (2006) observed that box size can affect the lying behaviour of horses; in their study they spent more time recumbent in a large box ([2.5× the height of the horse]^2 m^2) than in a small one ([1.5× the height of the horse]^2 m^2). An important observation raised by Pedersen and Ladewig (2004) is that single-boxed horses attempted to roll over before standing up. A possible explanation for this behaviour is an attempt to create distance from the box wall in order to be able to make the forward movement to get up. Rolling attempts can lead to different welfare problems, for example they can increase the risk of the horse becoming stuck against the box wall (ie becoming 'cast'), therefore lying space should be checked to ensure it is appropriate.

Although lying behaviour has never been tested for validity, reliability and repeatability, equine welfare scientists involved in the discussion agreed that this behaviour can be considered as a well-founded measure for assessing comfort around resting. Data are available on the time budgets for lying behaviour in horses; however measuring time budgets is very time consuming, therefore not truly feasible during a brief on-farm assessment.

Undoubtedly, there are a lack of animal-based indicators for assessing comfort around resting. In some cases, it may be helpful and easier to ask specific questions to the caretakers (eg what is your horse's preferred resting position?), even if it may lack objectivity. To address the Criterion comfort around resting, the absence of fresh/recent hock injuries, along with difficulties in getting up, should be considered as

https://doi.org/10.7120/09627286.23.3.323 Published online by Cambridge University Press

promising new animal-based indicators, as well as resource-based indicators, such as amount of lying space and quality of bedding. In some cases, as proven by Houpt and colleagues (2001), horses with no previous experience in straight stalls may be reluctant to lie down. In this study, nine of 16 mares kept in straight stalls were not observed in recumbency throughout a six-month observation period. Therefore, it should be borne in mind that, when insufficient lying space is provided (less than the small box measures reported by Raabymagle & Ladewig 2006), or the quality of bedding is very poor, horses do not lie at all, so neither getting up nor hock injuries can be observed. In this case resource-based measures are highly indicative of inadequate comfort.

Criterion: thermal comfort

This Criterion states that "animals should neither be too hot nor too cold" (Welfare Quality® Consortium 2009). The literature has largely focused on behavioural signs of heat stress in working equines in developing countries (Pritchard *et al* 2005, 2006; Minka & Ayo 2007; Burn *et al* 2009). Recently, Holcomb *et al* (2013) examined how behavioural and physiological parameters can be affected by hot temperatures in horses kept in on-farm environments and found that mature horses showed a preference for using shade in summer conditions; shade provided significant physiological benefits even with limited use. Increased frequency and depth of respiration, flared nostrils, profuse sweating, head nodding and apathy are behavioural signs used to assess the presence of heat stress. If four or more of these signs are observable on the same subject, the animal is suffering from heat stress (Pritchard *et al* 2005, 2006; Burn *et al* 2009). Heat stress is the only animal-based indicator that has been tested for all parameters and found to be valid, repeatable and reliable for assessing this criterion. On-farm feasibility was considered by different authors for assessing behavioural signs of heat stress (Pritchard *et al* 2005; Burn *et al* 2009; Holcomb *et al* 2013), in different housing and management conditions.

As well as heat, cold temperatures might affect the welfare of equines that do not have any shelter. Shivering, shelter seeking and huddling are assumed to be important behavioural signs of cold stress (Mejdell & Bøe 2005; Heleski & Murtazashvili 2010). Heleski and Murtazashvili (2010) studied shelter-seeking behaviour (SSB) and found that more horses used shelters in rainy, breezy conditions; in particular, the probability of SSB increased if the temperature was less than –1°C. Shivering is usually an acute response to a sudden cold exposure. Shivering and SSB could be considered valid measures of thermal comfort in cold temperatures (Mejdell & Bøe 2005; Heleski & Murtazashvili 2010). Although inter-observer reliability has not been evaluated for shivering and SSB, experts considered that good reliability could easily be achieved with training of assessors. Behavioural signs of cold stress also seem to be feasible and acceptable on-farm animal-based indicators.

Signs of heat and cold stress can be easily used on-farm to assess thermal comfort. As the absence of a shelter in presence of extreme temperatures can definitively compromise the ability of thermoregulation, resource-based indicators, such as presence of an appropriate shelter of adequate dimension, should be taken into consideration together with animal-based measures.

Criterion: ease of movement

This Criterion asserts that "animals should have enough space to be able to move around freely" (Welfare Quality® Consortium 2009). Locomotion plays a key role in horses, because it has both positive physical and mental effects on their health and because we take advantage of their ability to move when we use them. A common method for keeping horses in a domestic environment is single-box housing; therefore, a good animal-based indicator for evaluating when confinement compromises their welfare is needed. Daily activity and the presence of abnormal locomotory behaviours were found as animal-based indicators described in the literature. Daily activity can be electronically recorded using a motion sensor (Chaplin & Gretgrix 2010). Although data can be collected for an exact calculation of the daily activity of the subject, the use of an electronic device for 24 h is not seen as acceptable from the farmer's point of view, whereas the presence of abnormal behaviours (eg locomotor stereotypies such as box walking, weaving, fence pacing, pawing, box kicking) can be directly observed.

Locomotor stereotypies have been partially linked with insufficient activity, however validity has not been tested in experimental studies (McGreevy *et al* 1995; Heleski *et al* 2002; Bachmann *et al* 2003; Ninomiya *et al* 2007b). McGreevy and colleagues (1995) found that horses are less likely to develop abnormal behaviour if they spend more time out of the stable. To confound matters, locomotor stereotypies may indicate a previous welfare status versus the current welfare status.

Although repeatability and inter-observer reliability were not evaluated for either of these indicators, it is considered plausible that inter-observer reliability in recognising locomotor stereotypies or signs of their presence is achievable with training of assessors (eg videos).

The presence of locomotor stereotypies seems to be a feasible and acceptable on-farm animal-based indicator, however if considered alone without any other measure, it is not specific enough to assess the ability of horses to move around freely. Therefore, resource-based indicators regarding facilities (eg the possibility of going out to pasture), as well as the ratio between horse and box measures, together with a management-based indicator such as a questionnaire concerning the daily activity of the animals should be helpful in assessing the Criterion ease of movement.

Principle: good health

Animal-based indicators to assess the Principle good health, their validity, reliability and on-farm feasibility are reported in Table 4.

Criterion: absence of injuries

The animal-based indicators found in the literature were: the occurrence of hair discolouration, hairless patches, skin lesions, swollen joints/tendons, sensitive back and lameness (Leeb *et al* 2003; Pritchard *et al* 2005; Burn *et al* 2009; Neijenhuis *et al* 2011; Vervaecke *et al* 2011). These conditions might be linked with the presence of pain.

Table 4 Animal-based indicators for assessing the Principle good health.

Animal-based welfare indicators	Species	Housing/ category	Validity	Test-retest reliability	Inter-observer reliability	On-farm feasibility	References
Absence of injuries							
Hair discolouration	H	S	–	–	–	Yes	Vervaecke *et al* 2011
Hairless patches	H, D	W	–	Yes	Yes	Yes	Pritchard *et al* 2005; Burn *et al* 2009; Mekuria & Abebe 2010; Vervaecke *et al* 2011
Skin lesions	H, D	S, W	–	Yes	Yes	Yes	Leeb *et al* 2003; Pritchard *et al* 2005; Burn *et al* 2009, 2010; Mekuria & Abebe 2010; Neijenhuis *et al* 2011; Vervaecke *et al* 2011
Swollen joints/tendons	H, D	W	–	Yes	Yes	Yes	Pritchard *et al* 2005; Burn *et al* 2009, 2010; Mekuria & Abebe 2010
Lameness	H, D	S, W	–	Yes	Yes	–	Pritchard *et al* 2005; Burn *et al* 2009, 2010; Mekuria & Abebe 2010; Neijenhuis *et al* 2011; Viñuela-Fernández *et al* 2011
Sensitive back	H	S	–	–	–	–	Neijenhuis *et al* 2011; Asknes & Mejdell 2012
Absence of diseases							
Ectoparasites	H, D	W	–	–	Yes	Yes	Leeb *et al* 2003; Pritchard *et al* 2005; Burn *et al* 2009, 2010; Mekuria & Abebe 2010
Coat health	H, D	W	–	Yes	Yes	Yes	Leeb *et al* 2003; Pritchard *et al* 2005; Burn *et al* 2009, 2010; Mekuria & Abebe 2010
Faecal soiling	H, D	W	–	Yes	Yes	Yes	Leeb *et al* 2003; Pritchard *et al* 2005; Burn *et al* 2009, 2010; Mekuria & Abebe 2010
Abnormal breathing/dyspnoea	H	S	–	–	–	–	Leeb *et al* 2003; Couëtil & Hoffman 2007; Kutasi *et al* 2011
Cough	H	S	–	–	–	–	Kutasi *et al* 2011
Ocular discharge	H, D	W	–	Yes	Yes	Yes	Leeb *et al* 2003; Pritchard *et al* 2005; Burn *et al* 2009; Mekuria & Abebe 2010
Nasal discharge	H, D	W	–	–	–	–	Leeb *et al* 2003; Kutasi *et al* 2011
Mucous membrane colour	H, D	W	–	Yes	Yes	Yes	Pritchard *et al* 2005; Burn *et al* 2009; Mekuria & Abebe 2010
Limb/hoof associated abnormalities	H, D	W	–	Yes	Yes	Yes	Leeb *et al* 2003; Pritchard *et al* 2005; Burn *et al* 2009, 2010; Mekuria & Abebe 2010
Absence of pain induced by management procedures							
Pain-related behaviours	H, D	S	Yes	–	Yes	–	Taylor *et al* 2002; Ashley *et al* 2005; Dalla Costa *et al* 2010
Composite measures pain score	H	S	Yes	Yes	Yes	Yes	Bussières *et al* 2008; van Loon *et al* 2010; Graubner *et al* 2011
Horse Grimace Scale	H	S	Yes	–	Yes	Yes	Minero *et al* 2013; Dalla Costa *et al* 2014

H: Horse; D: Donkey; S: Single box; P: Paddock; W: Working equine.
Yes: tested and valid, reliable or feasible; No: tested and not valid, not reliable or not feasible; not tested.

https://doi.org/10.7120/09627286.23.3.323 Published online by Cambridge University Press

Hair discoloration is generally noted as unnatural patches of white hairs, presumably caused by inappropriate equipment (Vervaecke *et al* 2011); it indicates a lesion occurred in the past. Hairless patches are an area with hair loss and undamaged skin; whereas with a lesion, the skin is damaged either in the form of a scar, scab or wound (Pritchard *et al* 2005; Burn *et al* 2009). Hair discoloration, hairless patches and skin lesions are assessed by visual inspection of the animal's body and are recorded either on a presence/absence basis (Burn *et al* 2009; Mekuria & Abebe 2010) or a 3-point scale (Leeb *et al* 2003). Only lesions covering an area greater than 2 × 2 cm, 1 × 4 cm rectangle or 2.3-cm diameter on the body are recorded (Pritchard *et al* 2005; Mekuria & Abebe 2010). Their presence can be influenced by the type (eg ridden vs pack equines), quantity and intensity of work and by the type and quality of the equipment used, as well as the presence of disease (eg ectoparasites) or aggressive social interactions. Therefore, their location on the body (eg mouth corners, girth/belly, tail), number and severity should be recorded. Swollen joints/tendons are assessed by visual inspection of the flexor tendons and fetlock joints (Burn *et al* 2010) and scored either on a 3-point scale (Leeb *et al* 2003; Burn *et al* 2009), a presence/absence basis (Pritchard *et al* 2005) or as normal/swollen (Burn *et al* 2010). A sensitive back is assessed via palpation of the sides of the spine and evaluating the tension and/or sensitivity of the back muscles of the horse and is scored using by a 3- or 4-point scale (Neijenhuis *et al* 2011; Asknes & Mejdell 2012). Lameness is assessed by visual inspection of the subject during locomotion and is scored either by ticking a visual analogue scale (Viñuela-Fernández *et al* 2011), a 3- or 5-point scale (Neijenhuis *et al* 2011; Viñuela-Fernández *et al* 2011) or on a presence/absence basis (Pritchard *et al* 2005; Burn *et al* 2009, 2010).

None of the indicators found for this Criterion have been scientifically tested for validity, but the presence of, for example, lesions, should be considered if there is evidence that injuries have occurred.

Test-retest reliability has only been evaluated for skin lesions, swollen joints/tendons and lameness (Burn *et al* 2009). Inter-observer reliability has been tested and considered good for swollen joints/tendons (Pritchard *et al* 2005; Burn *et al* 2009). However, it has been found to be controversial for skin lesions, with Burn *et al* (2009) reporting low reliability, while Pritchard and colleagues (2005) reported it to be good. Burn *et al* (2010) suggest that their low inter-observer reliability was confounded by uncertainties among observers, due to unclear interpretation of low scale range when scoring indicators, as well as unbalanced prevalence of many indicators. Inter-observer reliability of the assessment of lameness has proven to be difficult to achieve, requiring extensive training and personal experience of the observer (Viñuela-Fernández *et al* 2011). The use of a very simple scoring system (yes/no) is reported to be helpful to achieve good reliability among assessors (Burn *et al* 2009). Simple, user-friendly scoring systems and proper training of assessors are necessary to improve reliability when recording skin lesions and lameness on-farm.

All of the reported indicators have been used for welfare assessment on working equines and have been described as easy to conduct under field conditions, requiring no expensive equipment (Leeb *et al* 2003; Burn *et al* 2009). All are designed to be practical, rapid, and to minimise handling and disruption to the animal's working routine (Pritchard *et al* 2005; Mekuria & Abebe 2010). Some doubts have been raised regarding on-farm feasibility of lameness and back sensitivity assessments, thus feasibility needs to be verified, and so too does acceptance by farmers. In order to adequately evaluate these indicators, it is essential to handle and move the horse out of its box. All other animal-based indicators outlined seem to be acceptable, possibly due to their ease of use and low disruption to work.

Criterion: absence of disease

The presence of disease can be determined through use of animal-based measures, which may infer its presence, rather than diagnose a particular disease. Several indicators that suggest an animal may be suffering from an underlying disease were found: depressed stance and presence of pain-related behaviours (see also *Criterion: absence of pain induced by management procedures*), the presence of ectoparasites, unhealthy coat, faecal soiling, cough, abnormal breathing/dyspnoea, ocular and nasal discharge, changes in mucous membrane colour (MMC) and limb/hoof-associated abnormalities (McDonnell 2002; Leeb *et al* 2003; Pritchard *et al* 2005; Burn *et al* 2009, 2010; Mekuria & Abebe 2010). All of these indicators are assessed by visual inspection. The presence of ectoparasites (eg flies, lice, ticks) has been scored on both a 3-point scale (Leeb *et al* 2003) and on a presence/absence basis (Pritchard *et al* 2005; Burn *et al* 2009; Mekuria & Abebe 2010).

The assessment of coat health is performed by examination of the hair on both sides of the animal's neck and recording whether there are any signs of alteration, eg matted, scabby or scruffy hair (Leeb *et al* 2003; Pritchard *et al* 2005; Burn *et al* 2009; Mekuria & Abebe 2010). Faecal soiling is assessed by inspecting the area inside the thighs and down back of the hocks and recording the presence (yes/no) of any amount of soiling with fresh or dried-on liquid faeces (Leeb *et al* 2003; Pritchard *et al* 2005; Mekuria & Abebe 2010), when present it is an indicator of diarrhoea. The presence of a cough, abnormal breathing/dyspnoea and/or nasal discharge can be signs of respiratory disease (Leeb *et al* 2003; Couëtil & Hoffman 2007; Kutasi *et al* 2011). To assess the presence of abnormal breathing/dyspnoea the observer should examine whether expiration is supported by the muscles of the trunk and whether the nostrils are dilated. Ocular discharge (or eye abnormalities) may be scored on a presence/absence basis (Burn *et al* 2010), or on a 3-point scale (Pritchard *et al* 2005) ranging from signs of mild discharge to signs of ocular pain, keratitis, uveitis and blindness. Nostrils should also be clean and free from discharge in healthy animals. Mucous membrane colour (MMC), is assessed by observation of the upper gum (eg pinkish in colour when normal, and ranging from pale, yellow, white, or purple in colour if abnormal) and scored as

Table 5 Animal-based indicators for assessing the Principle good behaviour.

Animal-based welfare indicators	Species	Housing/ category	Validity	Test-retest reliability	Inter-observer reliability	On-farm feasibility	References
Expression of social behaviour							
Isolation test	H	S, P	Yes	–	–	No	Lansade *et al* 2008a
Attraction test	H	S, P	No	–	–	No	Lansade *et al* 2008a
Vocalisations	H	S, P	Yes	–	–	–	Harewood & McGowan 2005; Lansade *et al* 2008a
Aggressive behaviours and related injuries	H	P	–	–	–	–	McDonnell 2002; Knubben *et al* 2008
Allo-grooming	H	P	–	–	–	–	Feh & de Mazières 1993; McDonnell 2002
Expression of other behaviours							
Stereotypies, crib-biting, wind sucking, weaving, box walking, head nodding, tongue playing, door knocking, wood chewing	H	S	–	–	–	Yes	Dierendonck & Goodwin 2005; Mills & Riezebos 2005; Wickens & Heleski 2010; Sarrafchi & Blokhuis 2013
Novel object test	H	S, P	Yes	Yes	–	Yes	Le Scolan *et al* 1997; Wolff *et al* 1997; Visser *et al* 2002; Momozawa *et al* 2003; Christensen *et al* 2008; Lansade *et al* 2008c; Górecka-Bruzda *et al* 2011; Leiner & Fendt 2011
Startling test	H	S	Yes	Yes	–	Yes	Visser *et al* 2001; Christensen *et al* 2008; Lansade *et al* 2008c; Górecka-Bruzda *et al* 2011
Novel arena	H	S, P	Yes	Yes	–	–	Le Scolan *et al* 1997; Wolff *et al* 1997; Seaman *et al* 2002; Lansade *et al* 2008c
Restraint and human fear test	H	S, P	Yes	Yes	–	–	Le Scolan *et al* 1997; Wolff *et al* 1997; Visser *et al* 2001; Górecka-Bruzda *et al* 2011
Positive emotional state							
Play and affiliative behaviours	H, D	P	–	–	–	–	Boissy *et al* 2007
Qualitative Behaviour Assessment	H	S	Yes	–	Yes	Yes	Napolitano *et al* 2008; Minero *et al* 2009; Fleming & Paisley 2013
Good human-animal relationship							
Approach test	H, D	W	Yes	Yes	Yes	–	Pritchard *et al* 2005; Burn *et al* 2010; Popescu & Diugan 2013
Walking down side	H, D	W	Yes	Yes	Yes	–	Pritchard *et al* 2005; Burn *et al* 2010; Popescu & Diugan 2013
Chin contact	H, D	W	Yes	Yes	Yes	–	Pritchard *et al* 2005; Burn *et al* 2010; Popescu & Diugan 2013
Voluntary animal approach	H	S, P	Yes	–	Yes	Yes	Søndergaard & Halekoh 2003; Maros *et al* 2010; Dalla Costa *et al* 2012
Forced human approach	H	S, P	Yes	–	Yes	Yes	Søndergaard & Halekoh 2003; Dalla Costa *et al* 2012

H: Horse; D: Donkey; S: Single box; P: Paddock; W: Working equine.
Yes: tested and valid, reliable or feasible; No: tested and not valid, not reliable or not feasible; not tested.

https://doi.org/10.7120/09627286.23.3.323 Published online by Cambridge University Press

normal/abnormal (Pritchard *et al* 2005; Burn *et al* 2009; Mekuria & Abebe 2010). Limb/hoof-associated abnormalities may cause gait abnormality (Ross & Dyson 2010), as well as signs of neglect (eg hoof lesions, overgrown) and can increase the risk of lameness. They are evaluated through observation of the subject whilst moving to determine whether any hoof-associated problems have caused mild or severe lameness and/or gait abnormality. Standard lameness and gait abnormalities were generally examined, where time allowed, over a 20-m trot-away before returning to the observer (Pritchard *et al* 2005).

The validity of these indicators has never been scientifically tested, but they are universally recognised as clinical signs linked with the presence of disease.

Coat health, faecal soiling, ocular discharge, MMC and limb/hoof abnormalities were evaluated by Burn *et al* (2009) and considered to be repeatable indicators. The presence of ectoparasites, coat health, faecal soiling, ocular discharge, MMC and limb/hoof-associated abnormalities were tested for inter-observer reliability by different authors. Burn *et al* (2010) found low reliability among observers; while, Pritchard *et al* (2005) successfully tested inter-observer agreement. Burn *et al* (2010) explained their low inter-observer reliability as probably being due to the homogeneity of the studied population and suggested that to increase this parameter, a more diverse equine population should be selected.

Most of the indicators were used in a 'simple' way to assess the presence of disease on working equines, so they can be considered feasible for an on-farm welfare assessment and acceptable from the farmer's point of view.

In view of all the observations reported above, most of the indicators have been found to be valid, reliable, feasible and observers can easily be trained. Therefore, they can be used on-farm to assess the Criterion absence of disease.

Criterion: absence of pain induced by management procedures

This Criterion considers that "animals should not suffer pain induced by inappropriate management, handling, slaughter, or surgical procedures (eg castration without anaesthesia and/or analgesia)" (Welfare Quality® Consortium 2009). Pain can be provoked by different conditions and can compromise equine welfare, therefore animal-based indicators are needed to identify pain and evaluate when appropriate pain-relief treatment is advisable. Indicators of pain described in the literature are the presence of pain-related behaviours and composite measure pain scores.

Pain-related behaviours (eg considerable restlessness, flank watching, reluctance to move, abnormal weight distribution, weight-shifting, lowered head carriage — not associated with sleep/dozing — fixed stare, dilated nostrils, clenched jaw) are considered to be valid animal-based indicators, as their presence is clearly linked with the presence of pain (see Ashley *et al* 2005 for a review; Taylor *et al* 2002; Olmos *et al* 2010; Dalla Costa *et al* 2012c).

Other indicators that can be used are composite measure pain scores (eg composite pain scale [CPS], post abdominal surgery pain assessment scale [PASPAS]), carried out through a brief observation of the subject (eg 5–10 min). Composite measure pain scores are a result of focusing not only on the presence of pain-related behaviours and changes in normal behaviour patterns (eg loss of appetite), but also on physiological parameters (eg heart rate, rectal temperature, respiratory rate). The CPS has been successfully applied by several authors following both surgical procedures (eg castration) or in the presence of injury and disease such as laminitis and colic (Bussières *et al* 2008; van Loon *et al* 2010; Graubner *et al* 2011), and its validity has been tested (Bussières *et al* 2008; van Loon *et al* 2010).

Both pain-related behaviours and the composite measure pain scores show good inter-observer reliability (Bussières *et al* 2008; van Loon *et al* 2010; Graubner *et al* 2011; Dalla Costa *et al* 2012b).

On-farm feasibility was not directly considered by the authors, but composite measure pain scores are primarily used for pain assessment in everyday practice by equine clinicians. Both indicators might be well accepted by the farmer as they only require observation of the subject. Composite measure pain scores require no more than 5 min per subject to record and can easily be used on stabled horses. They could, therefore, be considered feasible to measure on-farm. The use of pain-related behaviours as indicators may have some limitations: considering that equines, as prey animals, can mask obvious signs of pain in the presence of an unknown human — especially when the pain is mild — pain-related behaviours may be subtle and not overtly obvious.

Recently, a new approach to pain evaluation has been developed in other species utilising the assessment of facial expressions, incorporated into species-specific 'grimace scales' (Langford *et al* 2010; Sotocinal *et al* 2011; Keating *et al* 2012). Equines are very expressive animals and some facial changes (eg fixed stare, dilated nostrils, clenched jaw) are already described and commonly used to identify the presence of pain. Therefore, AWIN scientists focused their research on the development of the Horse Grimace Scale (HGS), with preliminary results suggesting HGS could be a promising pain indicator (Minero *et al* 2013; Dalla Costa *et al* 2014).

As many management procedures (eg castration) are performed when welfare assessors are not on-farm, the effects of surgery should be measured using questionnaires and analgesic drugs administered, to prevent horses and donkeys suffering from pain following these routine procedures.

Principle: good behaviour

Animal-based indicators to assess the Principle good behaviour, their validity, reliability and on-farm feasibility are reported in Table 5.

Criterion: expression of social behaviours

This Criterion considers that "animals should be able to express normal, non-harmful, social behaviours (eg grooming)" (Welfare Quality® Consortium 2009). Horses are highly social herd animals that prefer to live in a group; thus contact with other conspecifics plays an important role in their welfare. As horses are commonly stabled in single boxes, animal-based indicators are needed to evaluate

whether their need for social contact is fulfilled. To date, the search for indicators has focused on two main topics: i) tests performed to assess sociability and stress linked with separation from conspecifics (isolation test, attraction test, vocalisations); and ii) tests to address the quality of social contacts (eg kicks, bites and skin lesions, allo-grooming).

Isolation stress has been shown to significantly reduce welfare in many social species (eg cows: Boe & Faerevik 2003; rats: Patterson-Kane *et al* 2002; pigs: Pedersen *et al* 2002; rodents: Rault 2012). The isolation test is primarily designed to determine the presence of distress, resulting from short-term separation from conspecifics, without the possibility of joining or communicating with them, and to observe their reaction to isolation for 5 min (eg escape attempt, movements, alertness) (Lansade *et al* 2008a). The attraction test consists of isolating the test horse at one end of a corridor, with the opportunity to join familiar horses at the opposite end — the aim of this test is to assess the reaction to a social attraction (Lansade *et al* 2008a). Murray *et al* (2013) highlighted the importance of social contact in donkeys, demonstrating that pair-bonded individuals will seek out and preferentially 'choose' their companion when presented alongside a familiar or unfamiliar donkey. Vocalisations (eg neighing, whinny, bray) are proven to increase in frequency when horses and donkeys are stressed during separation from other conspecifics (Harewood & McGowan 2005; Lansade *et al* 2008a). An animal may live in a good overall social environment, yet still show signs of separation stress, indirectly reducing the efficacy of other positive welfare measures already in place. Therefore, entering a stable and hearing a lot of vocalisations could be a sign of separation stress. It is important to underline that vocalisations have never been tested on-farm to assess equine welfare. Aggressive behaviour, such as biting and kicking, is reported to be normal equine behaviour, which helps to create and maintain long-lasting dominance hierarchies (McDonnell 2002; Knubben *et al* 2008). As a consequence, a stable social group leads to the establishment of bonds and affiliative interactions among subjects, allowing allo-grooming to become more frequent, thus helping to alleviate social tension (Feh & de Mazières 1993; McDonnell 2002). The mixing of different herds or changes in group composition can result in elevated aggressive behaviour, with a higher occurrence of both biting- and kicking-related injuries (Knubben *et al* 2008). Thus, the occurrence of biting, kicking and related injuries (eg skin lesions) can be used as animal-based indicators to assess the stability of hierarchies and may also indicate insufficient resource availability and acquisition; eg quantity of hay provided/feeding density; how much space is allowed to avoid conflict near desired resources.

Lansade *et al* (2008a) tested the validity of both the isolation and the attraction tests: the isolation test is reported to be valid, whilst the attraction test was not. Vocalisations, kicks, bites, skin lesions and allo-grooming have never been tested for their validity.

Repeatability and inter-observer reliability have not been evaluated for any of the indicators described.

Concerns have been raised regarding the feasibility and the acceptance by farmers for the isolation and attraction tests, because they require a lot of time, handling and disruption to the animal's working routine and this is not compatible with a brief on-farm welfare assessment.

Although equine social behaviour is well studied, assessing this on-farm or in a welfare assessment context might not be feasible and needs further development. At present, no animal-based indicators are available to fully assess the Criterion expression of social behaviour, particularly in single-box housed equines. Therefore, resource and management-based measures (assessing the quantity and quality of social contact between horses) should be collected, as well as other promising indicators, such as bite and kick related injuries, vocalisations and allo-grooming.

Criterion: expression of other behaviours

This Criterion considers that "animals should be able to express other normal behaviours, ie it should be possible for them to express species-specific natural behaviours such as foraging" (Welfare Quality® Consortium 2009). Some features of the environment of the domestic horse could act as a potential stressor by limiting the ability to perform normal species-specific behaviour, restricting feeding or locomotion and imposing social isolation (McBride & Hemmings 2009). An environment which lacks stimuli and provides little to no possibility to express natural behaviour may be responsible for the development of abnormal behaviours (eg stereotypies) (Broom & Kennedy 1993; Hothersall & Casey 2012).

Stereotypic behaviour is described as "repetitive behaviour with no obvious goal and function" (Mason 1991) and has been linked to poor welfare and sub-optimal environments (Cooper & Mason 1998; Cooper & Albentosa 2005). Stereotypies are normally performed as a result of learned responses to environmental challenges or changes; signs may include crib-biting, wind sucking, weaving, box walking, head nodding, tongue playing, door kicking (Dierendonck & Goodwin 2005; Mills & Riezebos 2005; Wickens & Heleski 2010; Sarrafchi & Blokhuis 2013). Stereotypies can be used as animal-based indicators when directly observed or, indirectly, when evidence of their presence is detectable in the stable (eg damage to the wall, box-door or bedding) and/or on the horse (eg anti-cribbing collars). Stereotypies can become habit-forming, therefore, particularly during on-farm assessment, it may be unclear whether any observed stereotypies are representative of the current situation or a previous sub-optimal situation. On-farm feasibility and acceptance by farmers of the assessment of the presence of stereotypies has never been verified, but it does not seem impractical or time-consuming.

Horses are a prey species and as such it is their nature, in fear-eliciting situations, to show flight reactions which can be dangerous for both horse and handler. The presence of a threat in a horse's immediate environment,

https://doi.org/10.7120/09627286.23.3.323 Published online by Cambridge University Press

coupled with a fearful temperament, plays an important role in determining a long-term negative emotional state and over-reaction to fear-eliciting stimuli. These reactions can, in turn, cause harsh human responses that can affect the human-horse relationship and further jeopardise the animal's welfare. Therefore, finding appropriate indicators for assessing fearfulness in horses has important practical implications, not only for horse welfare, but also for human safety. Fear tests are experimental situations designed to evaluate fear responses: novel object tests (eg plastic tarpaulin), startle tests (eg opening an umbrella), novel arena tests or restraint and human fear tests have been used by different authors to assess behavioural responses to a fear-eliciting stimulus (Le Scolan *et al* 1997; Wolff *et al* 1997; Visser *et al* 2001, 2002; Seaman *et al* 2002; Momozawa *et al* 2003; Christensen *et al* 2008; Lansade *et al* 2008b; Górecka-Bruzda *et al* 2011; Leiner & Fendt 2011). Parameters recorded have included measuring frequency of behaviours (eg glances, sniffing, licking or nibbling), latency to approach the stimuli, flight distance, vocalisations (eg snorting, snuffling), defaecation during the event as well as physiological parameters, such as heart rate before and after the test. Predictive and concurrent validity for fearfulness tests (novel object, startling, novel arena and human fear tests) have been confirmed (Le Scolan *et al* 1997; Wolff *et al* 1997; Visser *et al* 2001, 2002; Seaman *et al* 2002; Momozawa *et al* 2003; Lansade *et al* 2008b; Leiner & Fendt 2011). In particular, Góreka-Bruzda *et al* (2011) found the most reliable indicator of a fearfulness trait was the time to approach the new stimulus and experimenter. The same results were also found by Visser *et al* (2001), Christensen *et al* (2005) and Henry *et al* (2005). Results were found to be valid (predictive, convergent and discriminant validity were all tested) and repeatable; however, inter-observer reliability was not evaluated. Moreover, the test was performed on only one breed (the Polish cold blood); therefore validation in other breeds may be necessary.

Although the animal-based indicators to determine fearfulness can be carried out and measured easily (Lansade *et al* 2008b), time constraints to actually conduct the tests during an on-farm assessment may hinder their efficacy, so that their use, undoubtedly relevant, might be limited to comprehensive welfare assessments.

In summary, stereotypic behaviour and fear tests are considered valid and reliable measures, which can be used as animal-based indicators for assessing the Criterion expression of other behaviours. As these indicators do not completely evaluate when this need for expression is required, the recording of other management-based measures (eg questionnaires that assess the possibility of foraging freely) should be integrated. Moreover, as fear and startle tests have the potential to cause short- and long-term welfare issues, their use needs careful consideration.

Criterion: positive emotional state

This Criterion focuses on the emotional state of animals, suggesting that "negative emotions such as fear, distress, frustration or apathy should be avoided, whereas positive emotions such as security or contentment should be promoted" (Welfare Quality® Consortium 2009). The potential to assess the positive emotions that animals may experience has aroused scientific interest over the past few years, with the realisation that animal well-being and welfare are not merely based on the absence of negative effects, but also the presence of positive effects (Boissy *et al* 2007).

No animal-based indicators to evaluate this Criterion in equines have been found in the literature to date; however, Boissy *et al* (2007) suggest that some behaviours are indicative of positive emotional states (eg play, affiliative behaviours). If we consider that horses are frequently stabled in single boxes, it is clear that these behaviours can be difficult to observe, although they may be useful when assessing horses kept in a group. Mendl and colleagues (2010) recently investigated cognitive bias in animals and developed tests to measure whether manipulations designed to alter affective states (eg living in an inappropriate environment) were linked to cognitive bias in the manner they predicted. Although these studies should be regarded as a significant development in animal welfare science and their validity is generally accepted, there is no doubt that the feasibility of cognitive bias tests during a relatively brief on-farm welfare assessment is limited.

A relatively new and promising animal-based measure is qualitative behavioural assessment (QBA), which characterises behaviour as expressive body language and uses subjective descriptive terms:

> to capture the animal's dynamic style of interaction with the environment by considering the animal as a whole, thus providing an insight into the animal's quality of life (Wemelsfelder 2007).

QBA requires a limited observation period (10–15 min) in which the assessor focuses on 'how' the animal expresses any given behaviour. Descriptors may be a fixed list of expressive or emotional terms of behaviour, or observers may generate their own descriptors (free choice profiling). They are then qualitatively scored on a Visual Analogue Scale of the intensity of the perceived expression of behaviour, for example how relaxed or agitated the animal is perceived to be (Wemelsfelder 2007).

Qualitative behavioural assessment has already been used by various authors to evaluate horse behaviour (Napolitano *et al* 2008; Minero *et al* 2009; Fleming & Paisley 2013), and results to date indicate that a meaningful relationship exists between QBA and quantitative measures (frequency and duration of behaviours, eg activity). These studies confirmed what was previously found for other farm animals, thus QBA is a biologically valid form of assessment (Rutherford *et al* 2012; Wemelsfelder 2012, 2007).

Furthermore, QBA was found to have high inter-observer reliability in other species, eg pigs (Wemelsfelder 2012). However, it should be highlighted, as with any other skill, adequate training must be undertaken by observers to ensure its efficacy. A possible down-side for on-farm use of QBA is that the result of the assessment is not immediate; in fact, it always requires some form of statistical analysis. Thus, efforts should be focused on finding an easy way to collect and analyse the data. A possible solution to this problem is the development of software which can store and analyse the data automatically as soon as they are uploaded.

Play and affiliative behaviours have not been validated, nor tested for repeatability or reliability, although it is emphasised that they are important and potential welfare indicators of positive emotional states based on the studies of farm animals (Boissy *et al* 2007). Ambiguity regarding play behaviour arises when it transforms into fighting and may be difficult to reliably measure without training.

Another important issue to address is on-farm feasibility, due to the potentially extensive observation time required. In fact, the standards for on-farm welfare assessments and information systems need to be simplified, with both resource and financial implications reduced. Although there is an evident lack of animal-based indicators to evaluate this criterion, the use of QBA could be a promising, quick, non-invasive and feasible on-farm measurement of positive emotional state, even if adequate validation and prior training of assessors is required.

Criterion: good human-animal relationship

The basis for this Criterion is that "animals should be handled well in all situations, ie handlers should promote good human-animal relationships" (Welfare Quality® Consortium 2009). In order to carry out common management and husbandry practices, horses and donkeys must be handled daily and their level of confidence with humans not only influences their performance and behaviour, but their fear reactions, which could have detrimental effects on both their own safety and that of humans.

Different human-animal relationship tests (eg voluntary approach test, forced approach test, walking down side test, chin contact) have been identified in the literature and can be used to assess this Criterion. These measures are reported to be appropriate for evaluating the human-animal relationship by assessing avoidance or friendliness towards a human (Søndergaard & Halekoh 2003; Pritchard *et al* 2005; Burn *et al* 2010; Maros *et al* 2010; Popescu & Diugan 2013). The tests were assessed in both working and on-farm environments.

In working environments (eg pack, driving, draught equines) the subject is restrained and the human approach tests are generally conducted in a series of steps. In the approach test, the assessor begins the test from a distance of 3 m from the equine and, at a normal pace, approaches the animal and records its reaction (eg the animal is friendly or turns away from the assessor) (Pritchard *et al* 2005; Burn *et al* 2010; Popescu & Diugan 2013). The assessor then performs the walk down side test — walking from head to rear, then returning along the opposite side, taking note of signs of attention or interest. When the subject is a donkey, they also look for signs of a 'tail-tuck' (Burn *et al* 2010; Popescu & Diugan 2013).

In working equines, the acceptance of chin contact in response to human contact has also given insight into the human-animal bond (Pritchard *et al* 2005; Burn *et al* 2010; Popescu & Diugan 2013). In an on-farm environment, the tests are usually performed where the subject is free to move in a paddock/arena. It should be noted that safety measures are paramount around free roaming horses to avoid any potentially aggressive incidents when unfamiliar people are performing behaviour tests. During the voluntary animal approach test (VAA), an unfamiliar person enters the paddock and walks to the middle of it; latency until a horse has its head within a distance of 1 m and latency until the horse touches the person is recorded (the maximum test time is 3 min) (Søndergaard & Halekoh 2003; Maros *et al* 2010). In the forced human approach test (FHA), an unfamiliar person approaches the horse slowly with approximately one step per s with hands hanging by the side. If the horse stands still when the person is within a 2-m range, the person slowly raised his hand and attempts to touch the neck of the horse, recording the reaction to be touched using a 4-point scale (Søndergaard & Halekoh 2003).

In working equines, the approach, walking down side and chin contact tests appear to be valid and repeatable measurements of human-animal relationships (Pritchard *et al* 2005; Burn *et al* 2009, 2010; Popescu & Diugan 2013). Inter-observer reliability seems to be moderate (Pritchard *et al* 2005; Burn *et al* 2009), but the use of a simple scoring system with clear definitions of scores and intensive training of assessors may help to improve this. However, because these tests require the farmer's involvement, this may be a problem for the on-farm environment where the farmer's time may be limited.

VAA seems to be a valid measurement (Maros *et al* 2010); whilst validity of FHA has not been assessed. Good inter-observer reliability for both tests was reported when assessing single stabled horses (Dalla Costa *et al* 2012a).

On-farm feasibility has been reported for both VAA and FHA (Søndergaard & Halekoh 2003) as they require a maximum of three minutes to conduct and minimal handling of the animals. For the same reasons, they both seem acceptable from the farmer's point of view.

None of the human-animal relationship tests described in the literature for any species studied are completely free from possible confounding factors. However, the avoidance distance and the voluntary approach tests were reported to be valid measures to assess the human-animal relationship (Waiblinger *et al* 2006). Human-animal relationship tests can be used as animal-based indicators to assess the Criterion good human-animal relationship. Further studies are required to evaluate VAA and FHA repeatability, as well as feasibility and acceptance by farmers for the avoidance distance, walking down side and chin contact tests.

https://doi.org/10.7120/09627286.23.3.323 Published online by Cambridge University Press

Animal welfare implications

In working equines, research findings, derived from the development and application of a scientifically sound welfare assessment protocol have already contributed to welfare improvements. Development of a similar protocol for the on-farm environment could enable the improvement of equine welfare in this area too.

Conclusion

As the initial step in achieving the goals set out in the European AWIN research project, this review aimed to identify possible valid, reliable and feasible animal-based indicators applicable for an on-farm welfare assessment of horses and donkeys. Validity is a concept also associated with sensitivity (the indicator's ability to identify positive results) and specificity (the indicator's ability to identify negative results) (EFSA Panel on Animal Health and Welfare 2012). However, both sensitivity and specificity for animal-based indicators have rarely been considered in welfare research. From the information reported above, the effort by researchers to find animal-based indicators which will assist the assessment of equine welfare in difficult situations, such as those of working equines, is evident. Although some aspects of horse welfare have been thoroughly investigated and indicators seem ready for on-farm use (eg absence of prolonger hunger, absence of injuries and diseases), others highlight the lack of scientific research, particularly in terms of validity and reliability. Further research should address the development of indicators for the Criteria: absence of prolonged thirst; comfort around resting; ease of movement; expression of social behaviour; and expression of other behaviours. A thorough evaluation of the validity and reliability of indicators such as signs of cold stress, QBA and human-animal relationship tests is needed, as well as other promising animal-based indicators such as the Horse Grimace Scale, which needs to be tested for on-farm feasibility. Consequently, AWIN research will focus on these important topics. A final, but no less important issue that deserves enhanced attention is the need for animal-based indicators to feasibly assess on-farm pain in donkeys: lack of research in this area is possibly a consequence of both a relatively lower interest in this species and our inability to interpret subtle changes in donkey behaviour.

Acknowledgements

The authors wish to thank the EU VII Framework programme (FP7-KBBE-2010-4) for financing the Animal Welfare Indicators (AWIN) project. We also acknowledge all the involved experts for their large contribution during the process of reviewing papers, refining the indicators' list and valuable support of the study; Kate Byrne for her extensive and professional revisions of language and structure and Sara Barbieri for help in the revision of the manuscript.

References

Acock AC 2008 *A Gentle Introduction to Stata*. Stata Press: College Station, USA

Ashley FH, Waterman-Pearson AE and Whay HR 2005 Behavioural assessment of pain in horses and donkeys: application to clinical practice and future studies. *Equine Veterinary Journal 37*: 565-575. http://dx.doi.org/10.2746/042516405775314826

Asknes F and Mejdell C 2012 Back soreness is common among 'healthy' riding horses. *Proceedings of the 46th Congress of the International Society for Applied Ethology* p 236. 31 July-4 August 2012, Vienna, Austria

Bachmann I, Audigé L and Stauffacher M 2003 Risk factors associated with behavioural disorders of crib-biting, weaving and box-walking in Swiss horses. *Equine Veterinary Journal 35*: 158-163. http://dx.doi.org/10.2746/0425164037761142l6

Blokhuis HJ, Veissier I, Miele M and Jones B 2010 The Welfare Quality® project and beyond: safeguarding farm animal well-being. *Acta Agriculturæ Scandinavica, Section A - Animal Science 60*: 129-140

Boe KE and Faerevik G 2003 Grouping and social preferences in calves, heifers and cows. *Applied Animal Behaviour 3*: 175-190. http://dx.doi.org/10.1016/S0168-1591(02)00217-4

Boissy A, Manteuffel G, Jensen MB, Moe RO, Spruijt B, Keeling LJ, Winckler C, Forkman B, Dimitrov I, Langbein J, Bakken M, Veissier I and Aubert A 2007 Assessment of positive emotions in animals to improve their welfare. *Physiology & Behavior 92*: 375-397. http://dx.doi.org/10.1016/j.physbeh.2007.02.003

Brambell FWR 1965 *Report of the Technical Committee to Enquire into the Welfare of Animals Kept under Intensive Livestock Husbandry Systems*. HM Stationery Office: London, UK

Broom DM 2011 A history of animal welfare science. *Acta Biotheoretica 59*: 121-137. http://dx.doi.org/10.1007/s10441-011-9123-3

Broom DM and Kennedy M 1993 Stereotypies in horses: their relevance to welfare and causation. *Equine Veterinary Education 5*: 2042-3292. http://dx.doi.org/10.1111/j.2042-3292.1993.tb01026.x

Burden F 2012 Practical feeding and condition scoring for donkeys and mules. *Equine Veterinary Education 24*: 589-596. http://dx.doi.org/10.1111/j.2042-3292.2011.00314.x

Burkholder WJ 2000 Assessment of the provision of optimal nutrition. *Journal of the American Veterinary Medical Association 217*: 650-654. http://dx.doi.org/10.2460/javma.2000.217.650

Burn C, Pritchard J and Whay H 2009 Observer reliability for working equine welfare assessment: problems with high prevalences of certain results. *Animal Welfare 18*: 177-187

Burn CC, Dennison TL and Whay HR 2010 Relationships between behaviour and health in working horses, donkeys, and mules in developing countries. *Applied Animal Behaviour Science 126*: 109-118. http://dx.doi.org/10.1016/j.applanim.2010.06.007

Bussières G, Jacques C, Lainay O, Beauchamp G, Leblond A, Cadoré J-L, Desmaizières L-M, Cuvelliez SG, Troncy E and Desmaizie L 2008 Development of a composite orthopaedic pain scale in horses. *Research in Veterinary Science 85*: 294-306. http://dx.doi.org/10.1016/j.rvsc.2007.10.011

Cappai MG, Picciau M and Pinna W 2013 An integrated approach towards the nutritional assessment of the Sardinian donkey: a tool for clinical nutritionists. *Italian Journal of Animal Science 12*: 12-15. http://dx.doi.org/10.4081/ijas.2013.e29

Carrol C and Huntington P 1988 Body condition scoring and weight estimation of horses. *Equine Veterinary Journal 20*: 41-45. http://dx.doi.org/10.1111/j.2042-3306.1988.tb01451.x

Chaplin SJ and Gretgrix L 2010 Effect of housing conditions on activity and lying behaviour of horses. *Animal 4*: 792-795. http://dx.doi.org/10.1017/S1751731109991704

Christensen JW, Keeling LJ and Nielsen BL 2005 Responses of horses to novel visual, olfactory and auditory stimuli. *Applied Animal Behaviour Science 93(1-2)*: 53-65. http://dx.doi.org/10.1016/j.applanim.2005.06.017

Christensen JW, Malmkvist J, Nielsen BL and Keeling LJ 2008 Effects of a calm companion on fear reactions in naïve test horses. *Equine Veterinary Journal 40*: 46-50. http://dx.doi.org/10.2746/042516408X245171

Cooper J and Albentosa M 2005 Behavioural adaptation in the domestic horse: potential role of apparently abnormal responses including stereotypic behaviour. *Livestock Production Science 92*: 177-218. http://dx.doi.org/10.1016/j.livprodsci.2004.11.017

Cooper J and Mason G 1998 The identification of abnormal behavior and behavioural problems in stabled horses and their relationship to horse welfare: a comparative review. *Equine Veterinary Journal Supplement 27*: 5-9

Couëtil L and Hoffman A 2007 Inflammatory airway disease of horses. *Journal of Veterinary Internal Medicine 21*: 356-361. http://dx.doi.org/10.1111/j.1939-1676.2007.tb02975.x

Dalla Costa E, Bonaita C, Pedretti S, Govoni E, Guzzeloni A, Canali E and Minero M 2012a Inter-observer reliability of three human-horse relationship tests. *Eighth International Society of Equitation Science Conference* p 94. 18-20 July 2012, Edinburgh, UK

Dalla Costa E, Minero M, Lebelt D, Stucke D, Canali E and Leach M 2014 Development of the Horse Grimace Scale (HGS) as a pain assessment tool in horses undergoing routine castration. *PloS One 9*: e92281. http://dx.doi.org/10.1371/ journal.pone.0092281

Dalla Costa E, Rabolini A, Scelsa A, Canali E and Minero M 2012c A study on inter-observer reliability of castration pain assessment in horses. *Second International Equine Science Meeting* p 52. 16-19 March 2012, Regensburg, Germany

Dalla Costa E, Rabolini A, Scelsa A, Ravasio G, Pecile A, Lazzaretti S, Canali E and Minero M 2012b Behavioural indicators of pain in horses undergoing surgical castration. *Proceedings of the 46th Congress of the International Society for Applied Ethology* p 235. 31 July-4 August 2012, Vienna, Austria

Dierendonck M and van Goodwin D 2005 Social contact in horses: implications for human-horse interactions. In: Francien Henriëtte de Jonge (ed) *The Human-Animal Relationship: Forever and a Day* pp 65-82. Ruud van den Bos: Assen, The Netherlands

EFSA Panel on Animal Health and Welfare (AHAW) 2012 Statement on the use of animal-based measures to assess the welfare of animals. *EFSA Journal 10*: 1-29

Ellis J and Hollands T 1998 Accuracy of different methods of estimating the weight of horses. *Veterinary Record 143*: 335-337. http://dx.doi.org/10.1136/vr.143.12.335

Ellis J and Hollands T 2002 Use of height-specific weigh tapes to estimate the bodyweight of horses. *Veterinary Record 3*: 632-635. http://dx.doi.org/10.1136/vr.150.20.632

European Horse Network 2010 *Key Figures*. http://www.europeanhorsenetwork.eu/index.php?page=horse-industry-in-europe

Faostat 2011 www.faostat.fao.org

Farm Animal Welfare Council 1979 http://www.fawc.org.uk/freedoms.htm

Feh C and de Mazières J 1993 Grooming at a preferred site reduces heart rate in horses. *Animal Behaviour 46*: 1191-1194. http://dx.doi.org/10.1006/anbe.1993.1309

Fleming P and Paisley C 2013 Application of Qualitative Behavioural Assessment to horses during an endurance ride. *Animal Behaviour 144*: 80-88. http://dx.doi.org/10.1016/j.applanim.2012.12.001

Fraser D 2001 The 'new perception' of animal agriculture: legless cows, featherless chickens, and a need for genuine analysis. *Journal of Animal Science 79*: 634-641

Górecka-Bruzda A, Jastrzębska E, Sosnowska Z, Jaworski Z, Jezierski T and Chruszczewski MH 2011 Reactivity to humans and fearfulness tests: field validation in Polish Cold Blood Horses. *Applied Animal Behaviour Science 133*: 207-215. http://dx.doi.org/10.1016/j.applanim.2011.05.011

Graubner C, Gerber V, Doherr M and Spadavecchia C 2011 Clinical application and reliability of a post abdominal surgery pain assessment scale (PASPAS) in horses. *Veterinary Journal 188*: 178-183. http://dx.doi.org/10.1016/j.tvjl.2010.04.029

Harewood E and McGowan C 2005 Behavioral and physiological responses to stabling in naive horses. *Journal of Equine Veterinary Science 25*: 164-170. http://dx.doi.org/10.1016/j.jevs.2005.03.008

Heleski CR and Murtazashvili I 2010 Daytime shelter-seeking behavior in domestic horses. *Journal of Veterinary Behavior: Clinical Applications and Research 5*: 276-282. http://dx.doi.org/10.1016/j.jveb.2010.01.003

Heleski CR, Shelle AC, Nielsen BD and Zanella AJ 2002 Influence of housing on weanling horse behavior and subsequent welfare. *Applied Animal Behaviour Science 78*: 291-302. http://dx.doi.org/10.1016/S0168-1591(02)00108-9

Henneke DR, Potter GD, Kreider JL and Yeates BF 1983 Relationship between condition score, physical measurements and body fat percentage in mares. *Equine Veterinary Journal 15*: 371-372. http://dx.doi.org/10.1111/j.2042-3306.1983.tb01826.x

Henry S, Hemery D, Richard MA and Hausberger M 2005 Human-mare relationships and behaviour of foals toward humans. *Applied Animal Behaviour Science 93*: 341-362. http://dx.doi.org/10.1016/j.applanim.2005.01.008

Holcomb K, Tucker C and Stull C 2013 Do horses benefit from provision of shade in hot, sunny weather, and do they prefer it? *9th International Equitation Science Conference* p 45. 17-20 July 2013, Newark, USA

Hothersall B and Casey R 2012 Undesired behaviour in horses: a review of their development, prevention, management and association with welfare. *Equine Veterinary Education 24*: 479-485. http://dx.doi.org/10.1111/j.2042-3292.2011.00296.x

Houpt K, Houpt T and Johnson J 2001 The effect of exercise deprivation on the behaviour and physiology of straight stall confined pregnant mares. *Animal Welfare 10*: 257-267

Kamphaus RW and Frick PJ 2005 *Clinical Assessment of Child and Adolescent Personality and Behavior, Second Edition*. Springer: New York, NY, USA

Keating SCJ, Thomas A, Flecknell P and Leach MC 2012 Evaluation of EMLA cream for preventing pain during tattooing of rabbits: changes in physiological, behavioural and facial expression responses. *PloS One 7*: e44437. http://dx.doi.org/10.1371/ journal.pone.0044437

https://doi.org/10.7120/09627286.23.3.323 Published online by Cambridge University Press

Knierim U and Winckler C 2009 On-farm welfare assessment in cattle: validity, reliability and feasibility issues and future perspectives with special regard to the Welfare Quality® approach. *Animal Welfare 18*: 451-458

Knubben JM, Furst A, Gygax L and Stauffacher M 2008 Bite and kick injuries in horses: prevalence, risk factors and prevention. *Equine Veterinary Journal 40*: 219-223. http://dx.doi.org/10.2746/042516408X253118

Kutasi O, Balogh N, Lajos Z, Nagy K and Szenci O 2011 Diagnostic approaches for the assessment of equine chronic pulmonary disorders. *Journal of Equine Veterinary Science 31*: 400-410. http://dx.doi.org/10.1016/j.jevs.2011.03.018

Langford DJ, Bailey AL, Chanda ML, Clarke SE, Drummond TE, Echols S, Glick S, Ingrao J, Klassen-Ross T, Lacroix-Fralish ML, Matsumiya L, Sorge RE, Sotocinal SG, Tabaka JM, Wong D, van den Maagdenberg AMJM, Ferrari MD, Craig KD and Mogil JS 2010 Coding of facial expressions of pain in the laboratory mouse. *Nature Methods 7*: 447-449. http://dx.doi.org/10.1038/nmeth.1455

Lansade L, Bouissou M and Erhard H 2008a Reactivity to isolation and association with conspecifics: a temperament trait stable across time and situations. *Applied Animal Behaviour Science 115*: 355-373. http://dx.doi.org/10.1016/j.applanim.2007.03.003

Lansade L, Bouissou M and Erhard HW 2008b Fearfulness in horses: a temperament trait stable across time and situations. *Applied Animal Behaviour Science 115*: 355-373. http://dx.doi.org/10.1016/j.applanim.2007.03.003

Leeb C, Henstridge C, Dewhurst K and Bazeley K 2003 Welfare assessment of working donkeys: assessment of the impact of an animal healthcare project in West Kenya. *Animal Welfare 12*: 689-694

Leiner L and Fendt M 2011 Behavioural fear and heart rate responses of horses after exposure to novel objects: effects of habituation. *Applied Animal Behaviour Science 131*: 104-109. http://dx.doi.org/10.1016/j.applanim.2011.02.004

Le Scolan N, Hausberger M and Wolff A 1997 Stability over situations in temperamental traits of horses as revealed by experimental and scoring approaches. *Behavioural Processes 41*: 257-266. http://dx.doi.org/10.1016/S0376-6357(97)00052-1

Maros K, Boross B and Kubinyi E 2010 Approach and follow behaviour: possible indicators of the human-horse relationship. *Interaction Studies 3*: 410-427. http://dx.doi.org/ 10.1075/is.11.3.05mar

Martin P and Bateson P 2007 *Measuring Behaviour: An Introductory Guide, Third Edition*. Cambridge University Press: Cambridge, USA. http://dx.doi.org/10.1017/CBO9780511810893

Mason G 1991 Stereotypies: a critical review. *Animal Behaviour 41*: 1015-1037. http://dx.doi.org/10.1016/S0003-3472(05)80640-2

Mason G and Mendl M 1993 Why is there no simple way of measuring animal welfare? *Animal Welfare* 2: 301-319

McBride S and Hemmings A 2009 A neurologic perspective of equine stereotypy. *Journal of Equine Veterinary Science 29*: 10-16. http://dx.doi.org/10.1016/j.jevs.2008.11.008

McDonnell S 2002 Behaviour of horses. In: Jensen P (ed) *The Ethology of Domestic Animals* pp 119-129. CABI Publishing: New York, USA. http://dx.doi.org/10.1079/9780851996028.0119

McGreevy PD, Cripps PJ, French NP, Green LE and Nicol C 1995 Management factors associated with stereotypic and redirected behaviour in the Thoroughbred horse. *Equine Veterinary Journal 27*: 86-91. http://dx.doi.org/10.1111/j.2042-3306.1995.tb03041.x

Mejdell C and Bøe K 2005 Responses to climatic variables of horses housed outdoors under Nordic winter conditions. *Canadian Journal of Animal Science 85*: 301-308. http://dx.doi.org/10.4141/A04-066

Mekuria S and Abebe R 2010 Observation on major welfare problems of equine in Meskan district, Southern Ethiopia. *Livestock Research for Rural Development 22 (Article #48)*. http://www.lrrd.org/lrrd22/3/meku22048.htm

Mendl M, Burman OHP and Paul ES 2010 An integrative and functional framework for the study of animal emotion and mood. *Proceedings of The Royal Society Biological Sciences 277*: 2895-2904. http://dx.doi.org/10.1098/rspb.2010.0303

Mills DS and Riezebos M 2005 The role of the image of a conspecific in the regulation of stereotypic head movements in the horse. *Applied Animal Behaviour Science 91*: 155-165. http://dx.doi.org/10.1016/j.applanim.2004.08.027

Minero M, Dalla Costa E, Lebelt D, Stucke D, Canali E and Leach M 2013 Development of a facial expressions pain scale in horses undergoing routine castration. *9th International Equitation Science Conference* p 36. 31 July-4 August 2012, Delaware, USA

Minero M, Tosi MV, Canali E and Wemelsfelder F 2009 Quantitative and qualitative assessment of the response of foals to the presence of an unfamiliar human. *Applied Animal Behaviour Science 116*: 74-81. http://dx.doi.org/10.1016/j.applanim.2008.07.001

Minka NS and Ayo JO 2007 Effects of shade provision on some physiological parameters, behavior and performance of pack donkeys (*Equinus asinus*) during the hot-dry season. *Journal of Equine Science 18*: 39-46. http://dx.doi.org/10.1294/jes.18.39

Momozawa Y, Ono T, Sato F, Kikusui T, Takeuchi Y, Mori Y and Kusunose R 2003 Assessment of equine temperament by a questionnaire survey to caretakers and evaluation of its reliability by simultaneous behavior test. *Applied Animal Behaviour Science 84*: 127-138. http://dx.doi.org/10.1016/j.applanim.2003.08.001

Murray LMA, Byrne K and D'Eath RB 2013 Pair-bonding and companion recognition in domestic donkeys, *Equus asinus*. *Applied Animal Behaviour Science 143*: 67-74. http://dx.doi.org/10.1016/j.applanim.2012.11.005

Napolitano F, De Rosa G, Braghieri A, Grasso F, Bordi A and Wemelsfelder F 2008 The qualitative assessment of responsiveness to environmental challenge in horses and ponies. *Applied Animal Behaviour Science 109*: 342-354. http://dx.doi.org/10.1016/j.applanim.2007.03.009

Neijenhuis F, De Graaf-Roelfsema E, Wesselink H, Van Reenen C and Visser EK 2011 Towards a welfare monitoring system for horses in the Netherlands: prevalence of several health matters. In: van Dierendonck M, de Cocq P and Visser E (eds), *7th International Equitation Science Conference: Equitation Science: Principles and Practice: Science at Work* p 80. 27-29 October 2011, Hooge Mierde, The Netherlands

Ninomiya S and Kusunose R 2004 Effects of feeding methods on eating frustration in stabled horses. *Animal Science Journal 75*: 465-469. http://dx.doi.org/10.1111/j.1740-0929.2004.00214.x

Ninomiya S, Sato S and Kusunose R 2007a A note on a behavioural indicator of satisfaction in stabled horses. *Applied Animal Behaviour Science 106*: 184-189. http://dx.doi.org/10.1016/j.applanim.2006.06.011

Ninomiya S, Sato S and Sugawara K 2007b Weaving in stabled horses and its relationship to other behavioural traits. *Applied Animal Behaviour Science 106*: 134-143. http://dx.doi.org/10.1016/j.applanim.2006.06.014

Olmos G, Burden F and Gregory N 2010 An innovative approach for better understanding the signs of pain in donkeys: the associations of pain-related pathology with clinical and behavioural indicators. *The 6th International Colloquium on Working Equids: Learning from Others* p 407. 29 November-2 December 2010, New Delhi, India

Patterson-Kane E, Hunt M and Harper D 2002 Rats demand social contact. *Animal Welfare 11*: 327-332

Pearson RA and Ouassat M 1996 Estimation of the liveweight and body condition of working donkeys in Morocco. *Veterinary Record 138*: 229-233. http://dx.doi.org/10.1136/vr.138.10.229

Pedersen GR and Ladewig J 2004 The influence of bedding on the time horses spend recumbent. *Journal of Equine Veterinary Science 24*: 153-158. http://dx.doi.org/10.1016/j.jevs.2004.03.013

Pedersen LJ, Jensen MB, Hansen SW, Munksgaard L, Ladewig J and Matthews L 2002 Social isolation affects the motivation to work for food and straw in pigs as measured by operant conditioning techniques. *Applied Animal Behaviour Science 77*: 295-309. http://dx.doi.org/10.1016/S0168-1591(02)00066-7

Pond W, Bazer F and Rollin B 2011 *Animal Welfare in Animal Agriculture: Husbandry, Stewardship, and Sustainability in Animal Production, First Edition.* CRC Press: Boca Raton, USA

Popescu S and Diugan E-A 2013 The relationship between behavioral and other welfare indicators of working horses. *Journal of Equine Veterinary Science 33*: 1-12. http://dx.doi.org/10.1016/j.jevs.2012.04.001

Pritchard J, Barr A and Whay H 2007 Repeatability of a skin tent test for dehydration in working horses and donkeys. *Animal Welfare 16*: 181-183

Pritchard J, Barr ARS and Whay HR 2006 Validity of a behavioural measure of heat stress and a skin tent test for dehydration in working horses and donkeys. *Equine Veterinary Journal 38*: 433-438. http://dx.doi.org/10.2746/042516406778400646

Pritchard J, Burn C, Barr A and Whay H 2008 Validity of indicators of dehydration in working horses: a longitudinal study of changes in skin tent duration, mucous membrane dryness and drinking behaviour. *Equine Veterinary Journal 40*: 558-564. http://dx.doi.org/10.2746/042516408X297462

Pritchard JC, Lindberg AC, Main DCJ and Whay HR 2005 Assessment of the welfare of working horses, mules and donkeys, using health and behaviour parameters. *Preventive Veterinary Medicine 69*: 265-283. http://dx.doi.org/10.1016/j.prevetmed.2005.02.002

Raabymagle P and Ladewig J 2006 Lying behavior in horses in relation to box size. *Journal of Equine Veterinary Science 25*: 502-504

Rault J-L 2012 Friends with benefits: social support and its relevance for farm animal welfare. *Applied Animal Behaviour Science 136*: 1-14. http://dx.doi.org/10.1016/j.applanim.2011.10.002

Reavell D 1999 Measuring and estimating the weight of horses with tapes, formulae and by visual assessment. *Equine Veterinary Education 11*: 314-317. http://dx.doi.org/10.1111/j.2042-3292.1999.tb01562.x

Ross M and Dyson S 2010 *Diagnosis and Management of Lameness in the Horse, Second Edition*. Elsevier Saunders: Missouri, USA

Rushen J, Butterworth A and Swanson JC 2011 Animal behavior and well-being symposium. Farm animal welfare assurance: science and application. *Journal of Animal Science* 89: 1219-1228. http://dx.doi.org/10.2527/jas.2010-3589

Rutherford KMD, Donald RD, Lawrence AB and Wemelsfelder F 2012 Qualitative Behavioural Assessment of emotionality in pigs. *Applied Animal Behaviour Science 139*: 218-224. http://dx.doi.org/10.1016/j.applanim.2012.04.004

Sarrafchi A and Blokhuis HJ 2013 Equine stereotypic behaviors: causation, occurrence, and prevention. *Journal of Veterinary Behavior: Clinical Applications and Research 8*: 1-9. http://dx.doi.org/10.1016/j.jveb.2013.04.068

Seaman S, Davidson H and Waran N 2002 How reliable is temperament assessment in the domestic horse (*Equus caballus*)? *Applied Animal Behaviour Science 78*: 175-191. http://dx.doi.org/10.1016/S0168-1591(02)00095-3

Søndergaard E and Halekoh U 2003 Young horses' reactions to humans in relation to handling and social environment. *Applied Animal Behaviour Science 84*: 265-280. http://dx.doi.org/10.1016/j.applanim.2003.08.011

Sotocinal SG, Sorge RE, Zaloum A, Tuttle AH, Martin LJ, Wieskopf JS, Mapplebeck JCS, Wei P, Zhan S, Zhang S, McDougall JJ, King OD and Mogil JS 2011 The Rat Grimace Scale: a partially automated method for quantifying pain in the laboratory rat via facial expressions. *Molecular Pain 7*: 55. http://dx.doi.org/10.1186/1744-8069-7-55

Taylor PM, Pascoe PJ and Mama KR 2002 Diagnosing and treating pain in the horse. Where are we today? *The Veterinary Clinics of North America. Equine Practice 18*: 1-19. http://dx.doi.org/10.1016/S0749-0739(02)00009-3

Van Loon J, Back W, Hellebrekers LJ, van Weeren PR, Loon JPAM and Van Weeren V 2010 Application of a composite pain scale to objectively monitor horses with somatic and visceral pain under hospital conditions. *Journal of Equine Veterinary Science 30*: 641-649. http://dx.doi.org/10.1016/j.jevs.2010.09.011

Vervaecke H, Boydens M, De Nil M and Laevens H 2011 Pilot study on the occurrence of pressure marks on the body and mouth lesions in riding horses in Flanders. In: van Dierendonck M, de Cocq P and Visser E (eds) *7th International Equitation Science Conference: Equitation Science: Principles and Practice: Science at Work* p 37. 27-29 October 2011, Hooge Mierde, The Netherlands

Viñuela-Fernández I, Jones E, Chase-Topping ME and Price J 2011 Comparison of subjective scoring systems used to evaluate equine laminitis. *Veterinary Journal 188*: 171-177

Visser EK, Reenen CG, Van Hopster H, Schilder MBH, Knaap JH, Barneveld A and Blokhuis HJ 2001 Quantifying aspects of young horses' temperament: consistency of behavioural variables. *Applied Animal Behaviour Science 74*: 241-258. http://dx.doi.org/10.1016/S0168-1591(01)00177-0

Visser EK, Reenen CG, Van Werf JTN, Van Der Schilder and MBH Knaap JH 2002 Heart rate and heart rate variability during a novel object test and a handling test in young horses. *Physiology & Behavior 76*: 289-296

https://doi.org/10.7120/09627286.23.3.323 Published online by Cambridge University Press

Waiblinger S, Boivin X, Pedersen V, Tosi M-V, Janczak AM, Visser EK and Jones RB 2006 Assessing the human-animal relationship in farmed species: a critical review. *Applied Animal Behaviour Science 101*: 185-242. http://dx.doi.org/10.1016/j.applanim.2006.02.001

Welfare Quality Consortium® 2009 *Welfare Quality® Assessment Protocol for Cattle*. Welfare Quality Consortium: Lelystad, The Netherlands

Wemelsfelder F 2007 How animals communicate quality of life: the qualitative assessment of behaviour. *Animal Welfare 16*: 25-31

Wemelsfelder F 2012 Assessing pig body language: agreement and consistency between pig farmers, veterinarians, and animal activists. *Journal of Animal Science 90*: 3652-3665

Wickens CL and Heleski CR 2010 Crib-biting behavior in horses: a review. *Applied Animal Behaviour Science 128*: 1-9

Wolff A, Hausberger M and Le Scolan N 1997 Experimental tests to assess emotionality in horses. *Behavioural Processes 40*: 209-221