Animal Welfare 2017, 26: 59-65
ISSN 0962-7286
doi: 10.7120/09627286.26.1.059

Developing a horse welfare assessment protocol

SM Viksten*†, EK Visser‡, S Nyman§ and HJ Blokhuis†

† Swedish University of Agricultural Sciences, Department of Animal Environment and Health, Box 7068, SE-75007 Uppsala, Sweden
‡ Horsonality, Skipper 3, 8456 JB De Knipe, The Netherlands
§ Wallby Säteri, Skirö, 574 96 Vetlanda, Sweden
* Contact for correspondence and requests for reprints: Sofie.Viksten@slu.se

Abstract

This paper describes the development and pilot-testing of a horse welfare assessment protocol (HWAP). The HWAP consists of the collective measurement of numerous factors considered likely to affect a horse's welfare and is thereby designed to provide a holistic score of its welfare status and to identify potential risk factors. The draft protocol contains 47 measures: 15 animal-based, 24 resource-based and eight management-based. It was tested in the autumn at two Swedish riding schools using a total of 37 horses of varying breed, gender and age. Each assessment was repeated after 16–25 days. The results showed that 66% (31/47) of the measures had over 85% repeatability between assessments. Results indicated occurrence of behavioural issues, eg aggression and avoidance, and potential risk factors, such as inadequate management routines and feeding regimes. Using the HWAP, the assessment of up to 22 horses could be carried out in one day. Changes were proposed to the draft protocol which included incorporating an ethogram to assess the human-animal relationship and assessing bit-related injuries. We propose that the protocol might: i) provide a firm basis for the welfare monitoring of horses; ii) identify important potential risk factors; iii) guide welfare improvement and management practices for horse owners and stable managers; and iv) contribute to the development of certification schemes for horse facilities.

Keywords: *animal welfare, assessment, equine, horse, protocol, Welfare Quality®*

Introduction

Welfare is a multi-dimensional concept. It comprises physical and mental health, and includes aspects such as comfort, absence of hunger, thirst, disease and fear (Blokhuis *et al* 2010), and not least the animals' own experience of their environment (Mellor *et al* 2009). Thus, assessing an animal's welfare status requires protocols that include a variety of effective and practical measures covering these different dimensions. The results of such a holistic assessment can identify welfare problems and potential risk factors thereby enabling animal owners and managers to adopt appropriate remedial measures. Traditional assessment protocols for horses, such as the Swedish protocol for official welfare controls (Jordbruksverket 2009) comprise numerous measures but the majority are resource-based and simply ensure that the environmental conditions and resources, such as box size, water availability and noise levels, comply with legal requirements. In the last decade, such protocols used on farm animals have been criticised for not assessing the actual welfare status of the animals themselves (Bracke *et al* 1999; Blokhuis *et al* 2003). Indeed, although the available resources are clearly relevant to welfare, how they relate to the animals' actual welfare status is not always clear. Factors such as management, husbandry methods and genetic background can have a profound influence on the relationship between the quality of a resource and actual achieved welfare (Blokhuis *et al* 2013). Consequently, there has been increased focus on implementing animal-based parameters (observed behaviour, physical condition, injuries, disease) and identifying related risk factors to safeguard and improve welfare. Thus, an effective assessment protocol, irrespective of species, should examine welfare from the animal's point of view, monitor changes over time and include management- and resource-based measures (Sørensen *et al* 2001). Some protocols already implement more animal-based measures in farm animals, notably the Welfare Quality® (WQ®) protocols for cattle, poultry and pigs (Welfare Quality® 2009a,b,c).

Horses are individually managed and used mainly for leisure and sport which makes some of the measures applied in farm animal protocols less suitable (eg pain from tail-docking) and may necessitate the inclusion of other measures (eg time in training). Horse protocols include the Australian Welfare Protocol (AHIC 2011), the Assessment Protocol for Horses (Wageningen UR 2012) and Minimum Standards of Horse Care in the State of California (Miller 2010). Visser *et al* (2014) specifically assessed the prevalence of health disorders and possible risk factors in Dutch horse stables and showed that a mix of animal-, resource- and management-based factors are required to enable the identification of causal factors. Based on the above study

Table 1 The selected measures in the HWAP categorised within the framework of the welfare principles and criteria of the Welfare Quality® approach (Blokhuis *et al* 2013).

Welfare principles	Welfare criteria	Measures in HWAP
Good feeding	1. Absence of prolonged hunger	Body condition score (BCS)
		Distance to next horse feeding point (roughage)
		Height of feed (concentrate trough and roughage)
		Time with available roughage
	2. Absence of prolonged thirst	Availability and cleanliness of water in stable and paddock
		Type, height, function and flow of drinker
Good housing	3. Comfort around resting	Housing size (group, box or tie-up stall)
		Noise level
	4. Thermal comfort	Thermal discomfort: sweating or shivering
		Sum of relative humidity (RH) and temperature (T) in stable
		Fresh air inlets in stable
	5. Ease of movement	Time in training
		Housing type (group, box or tie-up stall)
		Ceiling height
		Paddock size
		Time in paddock
		Access to pasture on grass
Good health	6. Absence of injuries	Lameness
		Hoof quality
		Farrier intervals
		Wounds
		Collisions or slipping when moving from stable to paddock
		Paddock surface quality
	7. Absence of disease	Roughage fed without water
		Order of feed (concentrates versus roughage)
		Coughing
		Signs of hampered breathing
		Ocular and nasal discharge
		Mould and condensation in stable
		Skin and coat condition
		Signs of scratching in tail and mane
	8. Absence of discomfort caused by use	Equipment chafing
		Rug cleanliness
Appropriate behaviour	9. Expression of social behaviour	Possibility of social contact
		Group size in paddock
	10. Expression of other behaviours	Occurrence of unwanted behaviour
		Enrichment for feed-seeking behaviour
	11. Good human-animal relationship	No measures included
	12. Positive emotional state	Visual horizon in stable

The title of the eighth criterion was altered from 'Absence of pain induced by management procedures' into 'Absence of discomfort caused by use'.

https://doi.org/10.7120/09627286.26.1.059 Published online by Cambridge University Press

and the Assessment Protocol for Horses (Wageningen UR 2012), we designed a pilot protocol with the potential to reliably and objectively assess horse welfare under Swedish conditions by providing an overview of welfare issues and possible causal factors ('risk factors') through a combination of animal-, resource- and management-based measures. The purpose of this study was to test the feasibility of the protocol *in situ*, establish whether signs of sub-optimal welfare were identified, and determine where further research is required. The horse welfare assessment protocol (HWAP) was developed using the WQ® criteria and principles of welfare as a framework to ensure a holistic approach embracing the various dimensions of welfare. Unlike the WQ® approach, no attempt was made in this study to integrate the various measures into scores for different welfare criteria or principles or to categorise the welfare quality of the stable. Such integration processes are necessary to calculate a farm overall welfare score which enables certification schemes.

Materials and methods

Composing the draft protocol

A literature review yielded a list of welfare measures in three categories: animal- (AB, measured on the animal itself), resource- (RB, the animal's environment or available resources) and management-based (MB, dependent on managers' decisions and management practices). All measures were discussed with experts (in Sweden and abroad) who had considerable experience of horses and welfare assessment: three veterinarians with different fields of expertise, one certified (by the Swedish Boards of Agriculture) equine physiotherapist, two certified (by the Swedish Boards of Agriculture) farriers, three animal scientists (PhDs) and two animal welfare officers working for the Swedish Government. Each measure was discussed in terms of: i) relevance to welfare; ii) feasibility *in situ*; iii) reliability; iv) the WQ® criterion in which it should be placed; and v) how to score it (scale and definition).

The title of the eighth WQ® criterion was, in accordance with Visser *et al* (2014), altered from 'Absence of pain induced by management procedures' to 'Absence of discomfort caused by use' since the original name refers to activities such as the dehorning of cattle which is irrelevant in horses.

After expert discussions, 47 non-invasive measures (15 AB, 24 RB, and eight MB) were considered relevant, potentially reliable and feasible for inclusion in the HWA protocol (see Table 1).

Stables, horses and assessments

The measures used here were pre-approved by the Uppsala Ethical Committee for Animal Testing (Permit reference number C145/11). All assessments were carried out by one person (SMV) who had previously worked for two years as an animal welfare officer in Sweden, had been handling horses on professional and recreational level for over 20 years and received training in lameness assessment, physical examinations and BCS scoring.

For safety reasons no naive horses were used, a number of horses were also excluded for other practical reasons (eg, veterinary visit or planned euthanasia between assessments). No brood mares, foals or stallions were included in this study. Assessments took place at two riding schools in mid-Sweden during October and November 2011. The riding schools were selected because they each had more than ten horses and personnel available to assist the assessor. In Stable 1, 15 of the 22 present horses and ponies were used in the study. They were of various breeds (Gotland pony, Swedish Warmblood, Swedish Coldblood and Norwegian Fjord horse), ages (4–23 years), gender (one mare and 14 geldings) and a mean (± SEM) withers height of 1.37 (± 0.2) m. Four horses (six at the second assessment) were individually stabled in boxes and eleven (nine at the second assessment) in tie-up stalls. The horses were ridden for 11.7 (± 7.9) h per week, the stable was approximately 30 years old with mechanical ventilation and a ceiling height of 2.71–3.0 m. In Stable 2, 22 of the 30 present horses were used in the study; these were of various breeds (Thoroughbreds, Swedish and Polish Warmbloods), ages (6–19 years), gender (four mares and 18 geldings) and a mean withers height of 1.64 (± 0.1) m. Twelve (13 at the second assessment) were housed in individual boxes and ten (nine at the second assessment) in tie-up stalls. The horses were ridden for 14.8 (± 4.1) h per week, the stable was less than ten years old with a computerised mechanical/natural ventilation system and a ceiling height of 2.96–8.4 m. All the horses in the study had bedding consisting of wood-shavings, were stabled at night and kept in groups of various sizes in paddocks during the day. Stable 1 kept horses in mixed gender groups and Stable 2 separated them according to gender in the paddocks.

All horses returned from pasture (complete rest and *ad libitum* access to grass) within a week prior to the first assessment.

Assessments began at 0600h before the horses were fed and let out into the paddocks.

The simplified lameness assessment (see Table 2[a]–[f] in the supplementary material to papers published in *Animal Welfare* on the UFAW website: http://www.ufaw.org.uk/t-ufaw-journal/supplementary-material) was novel to both stables that did not have a routine of assessment on a hard even surface.

Assessments were repeated in Stable 1 after 25 days and in Stable 2 after 16 days by the same assessor to test the reliability of each measure. The time-period between assessments was considered long enough to minimise risk of the assessor remembering previous results in the second assessment and short enough to minimise actual changes occurring. The assessor did not review or analyse results between assessments to further minimise risk of bias in results. Stable managers received no feedback between assessments.

Relative humidity (RH) and temperature (T) were recorded using a Relative Humidity and Temperature meter (Geo Fennel model FHT100, Geo-Fennel GmbH, Baunatal, Germany) outside before entering the stable and inside before the horses were taken out. All measures of length in the stables, feed heights, box size etc were recorded with a laser distance meter (Leica Disto model D2, Leica Geosystems AG, St Gallen, Switzerland).

Statistical analysis

All data were tested for normality using the Ryan-Joiner test. The results of each individual horse were compared between assessments and repeatability estimated using Cohen's kappa and Kendall's tau for ordinal data. Results from kappa analysis were interpreted using the Landis and Koch (1977) scale. Normally distributed data were analysed using a paired *t*-test. Coefficients of variance (CV, presented in square brackets) were used to analyse continuous data. Percentage agreement and standard deviation (SD) were calculated for applicable measures. All analyses were run at 5% significance level using a computer statistics package (Minitab® version 16.1.0, Minitab Ltd. UK).

Horses assessed in only one of the two assessments were excluded from analysis. The measures individual feeding amounts, height at withers, age, time of onset of undesirable behaviour and yearly access to pasture on grass were also excluded since they derived from information stated by the stable manager and did not change between assessments (see Table 2[f]; http://www.ufaw.org.uk/t-ufaw-journal/supplementary-material). Measures with 100% agreement between assessments were not analysed (see Table 3 in the supplementary material to papers published in *Animal Welfare* on the UFAW website: http://www.ufaw.org.uk/t-ufaw-journal/supplementary-material).

Results

The results show that 66% (31/47) of the measures incorporated in the HWAP had over 85% repeatability across two assessments carried out several days apart (see Table 3 and marked measures in Table 4 [supplementary material to papers published in *Animal Welfare* on the UFAW website: http://www.ufaw.org.uk/t-ufaw-journal/supplementary-material]). The time needed to carry out the actual assessment was 10–15 min per horse for animal-based measures, and the subsequent inclusion of all RB and MB measures meant that a stable could be assessed in one working day (less than or equal to 8 h). The assessment caused little disturbance to management routines.

Significant differences (poor agreement in results) between the first and second assessment were seen in BCS, distance to adjacent feeding point, height of concentrate troughs (in Stable 2), cleanliness of drinkers, sum of RH and T, open fresh-air inlets, lameness, paddock surface quality, mould, skin condition and equipment chafing (see Table 4; http://www.ufaw.org.uk/t-ufaw-journal/supplementary-material). The distributions of scores for measures rated on an ordinal scale were negatively skewed in all measures in either one or both assessments (see score distribution in Table 4; http://www.ufaw.org.uk/t-ufaw-journal/supplementary-material).

The RB measures taken with a laser meter showed different levels of variance in both assessments (mean [± SD] [CV]): height of concentrate troughs 0.8 (± 0.2) [19.9–20.2%]; height of drinkers in stable 0.9 (± 0.2) [21.9–26.0%]; housing size 7.4–7.8 (± 2.0) [27.2%] and ceiling height 4.9–5.0 (± 2.2) [44.2–44.3%].

Discussion

The HWAP had a high repeatability over two assessments in 66% of the incorporated measures. Moreover, it was shown that the protocol can be completed within one day for stables with up to 22 horses and may therefore be acceptable to horse owners and managers, particularly if future refinement shortens the duration.

Conversely, some measures failed the repeatability test. For example, the body condition scores (BCS) differed between first and second assessments at both stables. This methodology from Carrol and Huntington (1988) is validated and shown to be repeatable and the assessor was trained in it. Results therefore most likely reflect a real change caused by the fact that horses returned from pasture (complete rest and *ad libitum* access to grass) immediately prior to the first assessment and were then used in daily riding lessons, apparently reducing BCS in horses with a score above 3 (ie, less overweight). It is therefore still considered a relevant measure but, as with many others, it must be placed in the wider context of other potentially influential variables (eg, season, use of the horse, feeding regime) and thereby requires regular monitoring.

Measures of ventilation, RH, T and fresh-air inlets, also differed significantly between assessments. The manually regulated inlets were all closed in Stable 1 at both assessments so this precluded any significant change between assessments regardless of RH and T. Conversely, Stable 2 had automatically regulated inlets with RH and T sensors. However, the sum values of RH and T in both stables exceeded the recommended 80 or 90 (Ehrlemark 1994; CIGR 2012). This may reflect a lack of continuous regulation in Stable 1 and inappropriate settings of the climate system in Stable 2. Regulation of ventilation systems should account for variations in the indoor climate and ensure low sum values of RH and T. Regulation should be a factor to include in further studies and protocols. Moreover, mould was seen in both stables at the first assessment; this was unchanged in Stable 2 and reduced in Stable 1 between assessments probably because an annual clean took place during the interim. Mould and other endotoxins occur naturally in stable air and can cause airway disease (Wålinder *et al* 2011). Mould growth depends on air humidity (Nielsen 1979) so its occurrence may be an indicator of insufficient ventilation. Written feeding regimes stated that feed was given four times a day in both stables; this is normally sufficient to prevent long-term hunger and reduce health risks if adequate amounts and types of feed are provided and all horses in this study were assessed as having access to roughage (hay or hay silage) more than 6 h a day (Reeves *et al* 1996; Tinker *et al* 1997). However, the horses were not always allowed enough time to finish feeding before being taken for training or to the paddock. Both stables removed roughage from the stable when horses were in the paddock so the horses may not have eaten as much as stated in the written feeding regimes. This issue was not revealed by the current measures and illustrates the importance of including additional information regarding stable

https://doi.org/10.7120/09627286.26.1.059 Published online by Cambridge University Press

routines and daily schedules for feeding, training and rest. The fact that outcomes can change between assessments (depending on time between assessments) may reflect differences in prevailing conditions rather than unreliability of the measures. This can be expected for several RB and MB variables, eg quality of paddock surface (affected by weather), time in training (altered individually depending on a number of factors) etc. Such variables are often indicated as possible risk factors but their appropriate manipulation can help to improve horse welfare. Skin condition (AB) was also inconsistent over time but this can be related to diminished insect pressure as a result of season, being outdoors or indoors, to disease or skin disorders.

Measures such as occurrence of wounds, ocular discharge and drinker function showed high percentage agreement between assessments but had (where applicable) very low kappa values. This reflects the score distribution (towards 0) for these negatively skewed measures and the fact that some scores were missing. Both stables had high welfare standards so hardly any high scores were given and many measures were scored 0 at both assessments (ie 100% agreement). This makes it difficult to analyse reliability and indicates the need for a large sample size with a more heterogeneous population of horses (ie occurrence of all possible scores) and more distinct definition of scores. This sort of skewed scoring, where some scores are lacking or barely occurring, is common in welfare assessment (Kristensen *et al* 2006; D'Eath 2012). Percentage agreement is good to display reoccurrence of scores in an easy way but is fairly limited for data that varies very little (eg RB measures) where, for example, means and standard deviations along with coefficient of variance can say more about the nature of the results.

Measurements taken with a laser meter displayed a degree of variation as can be seen in Table 4 (http://www.ufaw.org.uk/t-ufaw-journal/supplementary-material). The highest variation was in ceiling height which was measured from the top of the bedding to the ceiling. The higher variation was due to bedding thickness variation across individual stables. Since it is likely that the instrument measures height correctly this problem could be solved through measuring from the floor as opposed to the top of the bedding. Shivering and sweating were not apparent but such signs of thermal discomfort might only be expected, and be appropriate as a measure, during more demanding weather conditions in midsummer or midwinter.

Several horses shied away from adjacent ones whilst feeding after visual threats (bared teeth and pinned-back ears), this was more frequent in tie-up stalls than in boxes. This observation is in line with previous studies showing that visual contact with higher ranked horses can hamper feeding in lower rank horses (Krüger & Flauger 2008), therefore behaviours such as shying away during feeding should be included in the HWAP if feasible.

None of the stables assessed lameness as a daily routine. By including a simplified lameness assessment, such as the one in HWAP, early signs of lameness can be picked up and treatment started at an early stage.

The absence of skin disorders on legs and pasterns at both stables despite some high scores on paddock surface indicates that even if paddock surface is a risk-factor, proper management and preventative measures can prevent welfare issues occurring. It might also indicate that there were enough dry patches in the paddock for horses to stand on to ensure the more muddy parts did not affect horses as much (Akucewich & Anthony 2007; Scott & Miller 2011). The low frequency of injuries and wounds could be partly explained by good management (eg horses were seldom regrouped [Christensen *et al* 2011]). Risks of injury were further reduced because the stables and paddocks were kept clear of obstacles and other items that could cause injury which again emphasises the need to include information regarding managerial routines in welfare assessments.

Assessment of wounds can be misleading since equipment chafing can also display as wounds in certain areas, such as the corners of the mouth. A large-scale welfare assessment in The Netherlands revealed that over 3% of horses had open wounds or irregularities in the corners of the mouth (Visser *et al* 2014). Therefore, the location of the wound should be recorded. Such data could inform the future fitting (and perchance design) of equipment.

The assessment of undesirable behaviour (eg stereotypy and aggression) took place around feeding time and before the horses went outside. This is when frequency was expected to peak (Cooper & Mason 1998). Some horses were aggressive towards the assessor and other horses in both stables and had to be held by the handler during the animal-based measures.

Limitations of the pilot study

The stables participated voluntarily and it may be that only those stables already achieving a high level of welfare agreed to be involved, which might be a selection bias. Having high standards in regard to welfare may result in skewed outcomes, which was the case in this study. The sample size in the current study was relatively small which meant that the statistical analyses were occasionally affected. The availability of only one assessor excluded testing for inter-observer repeatability and is a possible source of error in regards to the assessor remembering previous results from the first assessment. This risk was reduced however by blinding the assessor to previous results (not accessing them or analysing them) between assessments. The time between assessments was a compromise between minimising the risk mentioned (requiring a long interval) and preventing the risk of real changes that could occur as a result of changes in the weather or injuries to the horses. In order to evaluate if the HWAP is valid for monitoring changes over long periods, a longer study is required with several repeated assessments.

Although the horses primarily included geldings, we do not know of any studies implying that this might affect results. Not all horses present at the stables were used in all measures or even in both assessments since a number were indisposable due to being away at training, being sold (not present at stable anymore) or euthanased between assessments due to injuries unrelated to the study. This reflects the challenges associated with doing repeated assessments in the field: the sample population does not always remain constant.

Suggested ways of improving the HWAP

The scoring of all measures should be harmonised to a scale of 0–2 in accordance with the WQ® system. This simplifies statistical analyses and can be used when calculating a farm's overall score. This applies to measures such as visual horizon that should be scored 0 = 3–4 sides, 1 = 1–2 sides and 2 = 0 sides.

The protocol should include a measure of the human-animal relationship (HAR) since this is currently lacking. A simplified ethogram could be added where behaviour towards the assessor (HAR) is scored as 0 = positive (seeking contact, ears forward), 1 = neutral (little interest in assessor) and 2 = aggressive or avoiding (threatening to bite or kick, moving away).

Identification of the best type of bedding and the most appropriate thickness would improve horses' comfort when lying and reduce the incidence of wounds.

Measurement of all heights of troughs, ceilings and water troughs should be conducted from the floor to reduce risks of the bedding thickness affecting results.

The measure 'distance from one feeding station to the next' should be altered to 'undisturbed feeding in regards to visual contact with other horses' and scored as 0 = yes and 1 = no. This is determined by assessing whether the partition walls allow horses to see each other when feeding.

There was concern that some horses were led away before they had finished feeding. The time required for a horse to either consume its entire roughage ration or voluntarily move away requires further study.

The number of horses per drinking point in the paddock should be included because high-ranking horses were observed guarding the troughs and deterring others from drinking.

Several horses displayed aggression or avoidance when being touched on the back during the physical assessment, in agreement with the study by Visser *et al* (2014). This might be pain-related so palpation of the back muscles should be added to enable early detection and onset of treatment to prevent behavioural change. Scored 0 = no or a small initial reaction that did not occur at a second attempt and 1 = visible muscle tension and abrupt lowering of back.

To further assess equipment and usage effects, the occurrence of bit-related injuries in the lower part of the mouth (Visser *et al* 2014) should be scored as 0 = no depigmentation or sign of chafing, 1 = depigmentation or chafing, 2 = open wounds or depigmentation and chafing.

Animal welfare implications

The main purpose of the horse welfare assessment protocol developed in this pilot study was to enable a standardised detection of possible welfare issues related to management regimes, housing conditions and use of the horse and to trigger further analysis of causes and remedial actions. Feedback of the assessment results to owners and managers will raise awareness and benefit horse welfare by enabling informed improvements in housing and management.

Acknowledgements

The authors are particularly grateful to all the stable owners and their personnel who made us welcome. We thank Dr Jan Hultgren for valuable advice on statistical analysis and Dr Peta Hitchens and Dr Elke Hartmann for helping us improve the manuscript. We also thank Dr Bryan Jones for valuable comments on the manuscript and improving the English language. This study was carried out within the Centre of Excellence in Animal Welfare Science, a Swedish collaborative research environment.

References

AHIC 2011 *Australian Horse Welfare Protocol*. Australian Horse Industry Council: Geelong, Victoria, Australia

Akucewich LH and Anthony AY 2007 Equine pastern dermatitis. *Compendium: Equine Edition, Volume 2, No 4*. http://www.vetfolio.com/search?q=equine+pastern+dermatitis

Blokhuis HJ, Jones RB, Geers R, Miele M and Veissier I 2003 Measuring and monitoring animal welfare: Transparency in the food product quality chain. *Animal Welfare 12*: 445-455

Blokhuis HJ, Jones RB, Veissier I and Miele M 2013 The Welfare Quality® vision, In: Blokhuis HJ, Jones RB, Veissier I and Miele M (eds) *Improving Farm Animal Welfare. Science and Society Working Together: The Welfare Quality Approach* pp 71-89. Wageningen Academic Publishers: Wageningen, The Netherlands. https://doi.org/10.3920/978-90-8686-770-7_4

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

Bracke MBM, Spruijt BM and Metz JHM 1999 Overall animal welfare assessment reviewed. Part 1: Is it possible? *NJAS Wageningen Journal of Life Sciences 47*: 279-291

Carroll CL and Huntington PJ 1988 Body condition scoring and weight estimation of horses. *Equine Veterinary Journal 20*: 41-45. https://doi.org/10.1111/j.2042-3306.1988.tb01451.x

Christensen JW, Søndergaard E, Thodberg K and Halekoh U 2011 Effects of repeated regrouping on horse behaviour and injuries. *Applied Animal Behaviour Science 133*: 199-206. https://doi.org/10.1016/j.applanim.2011.05.013

CIGR 2012 *International Commission of Agricultural and Biosystems Engineering. Working group reports; Climatization of Animal Houses.* http://cigr.org/Resources/working-group-reports.php

Cooper JJ and Mason GJ 1998 The identification of abnormal behaviour and behavioural problems in stabled horses and their relationship to horse welfare: a comparative review. *Equine Veterinary Journal 30*: 5-9. https://doi.org/10.1111/j.2042-3306.1998.tb05136.x

D'Eath RB 2012 Repeated locomotion scoring of a sow herd to measure lameness: consistency over time, the effect of sow characteristics and inter-observer reliability. *Animal Welfare 21*: 219-231. https://doi.org/10.7120/09627286.21.2.219

Ehrlemark A 1994 Naturlig ventilation i djurstallar. *Fakta-Teknik, nr 12*. Swedish University of Agricultural Sciences: Sweden

Jordbruksverket 2009 *Checklista häst version 1.0*. Swedish Board of Agriculture: Jönköping, Sweden

Kristensen E, Dueholm L, Vink D, Andersen J, Jakobsen E, Illum-Nielsen S, Petersen F and Enevoldsen C 2006 Within- and across-person uniformity of body condition scoring in Danish Holstein cattle. *Journal of Dairy Science 89*: 3721-3728. https://doi.org/10.3168/jds.S0022-0302(06)72413-4

Krüger K and Flauger B 2008 Social feeding decisions in horses (*Equus caballus*). *Behavioural Processes 78*: 76-83. https://doi.org/10.1016/j.beproc.2008.01.009

Landis J and Koch G 1977 The measurement of observer agreement for categorical data. *Biometrics 33*: 159-174. https://doi.org/10.2307/2529310

Mellor D, Patterson-Kane E and Stafford KJ 2009 *The Sciences of Animal Welfare*. John Wiley & Sons: West Sussex, UK

Miller G, Stull C, Ferraro G and Meierhenry B 2010 Minimum standards of horse care in the state of California. *Reserve for law enforcement of applicable federal and state regulations, codes and laws Edition*. Center of Equine Health: School of Veterinary Medicine, University of California, USA

Nielsen KF 1979 *Mould growth on building materials*. PhD Thesis, By og Byg, Statens Byggeforskningsinstitut, Norway

Reeves MJ, Salman MD and Smith G 1996 Risk factors for equine acute abdominal disease (colic): Results from a multi-center case-control study. *Preventive Veterinary Medicine 26*: 285-301. https://doi.org/10.1016/0167-5877(95)00551-X

Scott DW and Miller WH 2011 *Equine Dermatology*. Elsevier: Amsterdam, The Netherlands

Sørensen J, Sandøe P and Halberg N 2001 Animal welfare as one amongst several values to be considered at farm level: the idea of an ethical account for livestock farming. *Acta Agriculturae Scandinavica, Section A - Animal Science 30*: 11-16

Tinker MK, White NA, Lessard P, Thatcher CD, Pelzer KD, Davis B and Carmel DK 1997 Prospective study of equine colic risk factors. *Equine veterinary journal 29*: 454-458. https://doi.org/10.1111/j.2042-3306.1997.tb03158.x

Visser EK, de Graaf-Roelfsema E, Wesselink HGM, de Boer J, van Wijhe-Kiezebrink MC, Engel B and van Reenen CG 2014 Risk factors associated with health disorders in sport and leisure horses in The Netherlands. *Journal of Animal Science 92*: 844-855. https://doi.org/10.2527/jas.2013-6692

Wageningen UR 2012 *Welfare Monitoring System – Assessment Protocol for Horses, Version 2.0*. Wageningen UR Livestock Research: Wageningen, The Netherlands

Wålinder R, Riihimäki M, Bohlin S, Hogstedt C, Nordquist T, Raine A, Pringle J and Elfman L 2011 Installation of mechanical ventilation in a horse stable: effects on air quality and human and equine airways. *Environmental Health and Preventive Medicine 16*: 264-272. https://doi.org/10.1007/s12199-010-0195-5

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

Welfare Quality® 2009b *Welfare Quality® Assessment Protocol for Pigs*. Welfare Quality®: Lelystad, The Netherlands

Welfare Quality® 2009c *Welfare Quality® Assessment Protocol for Poultry*. Welfare Quality®: Lelystad, The Netherlands