University of
Natural Resources and
Life Sciences, Vienna
Department for
Sustainable Agriculture Systems
Division of Livestock Sciences

University of Natural Resources and Life Sciences

Department of Sustainable Agricultural Systems

Division of Livestock Science

Development and Testing of Parameters for the Hands-on Donkey Welfare Assessment Protocol

Master thesis

Submitted by

Mag.[a] Hanna Zedlacher BSc

Supervisors:

Prof. Dr. Christoph Winckler

Dr. Christine Leeb

Vienna

May 2014

Content

Figures

Tables

Photos

Abbreviations

BCS	Body Condition Score
BI	Bias index
DS	The Donkey Sanctuary
IOR	Inter-Observer Reliability
PCA	Principal Component Analysis
PI	Prevalence index
QBA	Qualitative Behaviour Assessment
WEWA	Working Equine Welfare Assessment
WQ®	Welfare Quality®

1 Introduction

Working equines play a very important role in many parts of the world. Whereas in most parts of Europe the population of working equines has declined since the end of the Second World War donkeys are still widely used in Africa, Asia and South America as pack, draft and riding animals. A high percentage of these animals – used in different working environments such as agriculture or industry (e.g. pack donkeys in brick kilns) – are suffering from various health problems and welfare issues (e.g. lameness, wounds).

Animal welfare charities such as *The Donkey Sanctuary* (www.thedonkeysanctuary.org.uk) or *The Brooke* (www.thebrooke.org) have been working in Africa, Asia and South America for several decades. These organizations provide veterinary treatment and offer training programmes or services (e.g. farriery, harness making) in various regions. Generally the approach of the majority of such charities is two sided: on the one hand they offer short term interventions such as veterinary treatment, on the other hand the focus also lies on preventing problems to occur in the first place. Here the provision of trainings etc. plays an important role.

The aim of all charities is improved welfare and health of working equines. Animal welfare assessment is therefore relevant in order to identify welfare issues and potential working areas or to test the effectiveness of interventions. For many years - not only in the field of working equines but also regarding farm animals in Europe - the effectiveness was evaluated by using input-based parameters (e.g. number of vaccines used, number of people taking part in trainings). A paradigm shift lead to a different focus: today additionally animal-based indicators are used to assess the welfare of animals and therefore the effect of interventions. However, the conditions for carrying out a welfare assessment of working equines differ from those on European farms.

The UK based animal welfare charity *The Donkey Sanctuary* (DS) works in several countries around the world. To be able to assess the welfare of donkeys in these regions, to monitor the charity's own progress and to identify potential new operating areas the DS created its own assessment protocol named Hands-on Donkey Welfare (Figure 1) (BLAKEWAY, 2013 personal communication) However there are no handbooks or (detailed) guidelines to carry out the welfare assessment. The five fingers represent the most important welfare issues or working areas of the DS. It has been used in a qualitative, subjective way which included an individual, subjective estimation of the worst and best place to live for a donkey with regard to the five fingers. The thumb stands for "Behaviour/Demeanour" whereas the index finger deals with the "Body Condition" (Body Condition Score, BCS) of the donkeys. The middle and the ring finger represent "Wounds" and "Lameness" respectively. The little finger covers the topic of "Other Signs of Injury and Disease". Within the framework of this Master thesis this existing protocol was extended and tested in Ethiopia in summer 2013 in order to apply it in a more structured, in-depth approach in the future.

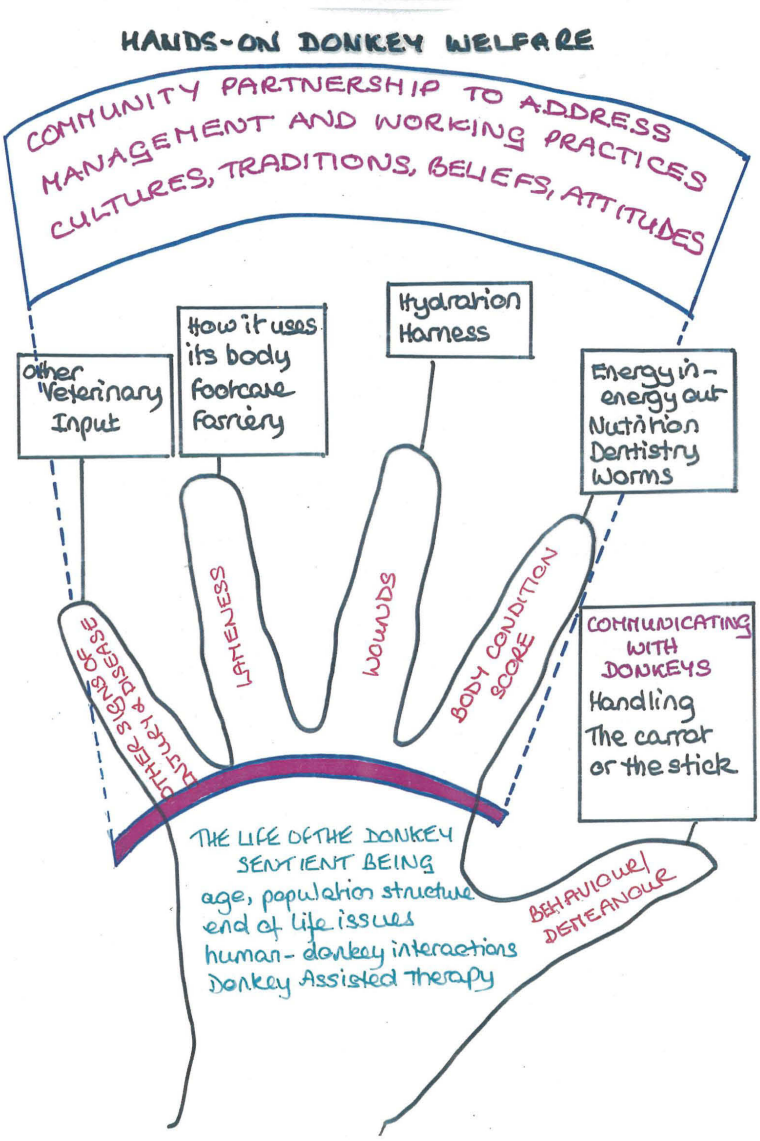


Figure 1: Hands-on Donkey Welfare

2 Aims and research questions

The aim of the present thesis is the collation and development of parameters which can be used in a more detailed Hands-on Donkey Welfare assessment protocol. Furthermore an on-field application of those parameters which were chosen to represent the five fingers (Figure 1) was carried out to test their feasibility. Furthermore their reliability during a donkey welfare assessment was tested by examining the consistency of ratings by different observers. Furthermore, Qualitative Behaviour Assessment (QBA) in donkeys as a pilot test is part of the present thesis. As QBA has never been carried out with donkeys before, interesting additional information was expected to be included in the final protocol. The overarching analysis of the obtained results aims for providing recommendations for a future application of the extended Hands-on Donkey Welfare assessment protocol.

In order to assess the welfare of working donkeys in a standardised, objective way several aspects have to be considered. Validity and practicability in the field are two major aspects in this matter. Furthermore the reliability of chosen parameters within a welfare assessment protocol is an important issue. These mentioned aspects are crucial; therefore the challenges and benefits of the chosen approach will be tested and discussed.

Research questions:

- What are appropriate (i.e. valid, reliable and accepted by the local stakeholders) parameters for the five “fingers”? To what extent is Qualitative Behaviour Assessment (QBA) a useful parameter to assess the welfare of working donkeys in Ethiopia?
- Which aspects/parameters regarding donkey welfare are considered by the DS staff as important to assess?
- How consistent are the ratings between assessors which have taken part in the development of the detailed protocol?
- What is the state of donkey welfare in selected regions using the newly developed “extended” Hands-on Donkey Welfare assessment protocol?

3 Current state of research

The following chapters deal with concepts of animal welfare as well as animal welfare assessment. Additionally, a literature overview of donkey welfare describes the main welfare issues with a strong focus on developing countries.

3.1 Animal Welfare Concepts

Three commonly used scientific concepts of animal welfare will be described briefly in the following paragraph.

The first concept defines animal welfare as the attempt to cope with the environment and the failure to cope with the environment respectively. BROOM (1996) describes the concept as "states as the origin of the concept "how well the individual is faring or travelling through life". The physiological balance must be maintained as well as the animal's needs must be fulfilled in order to safeguard welfare. Examples for failures to cope with the environment are the occurrence of diseases and injuries as well as physiological or pathological alterations (e.g. lesions in the lungs of veal calves, foot pad dermatitis in broilers). Additionally, physiological or clinical signs of stress may serve as an indicator for poor welfare as the animal is not able to cope with the environment. Other experts do not see stress per se as a factor for reduced welfare as it occurs in an animal's life without being fundamentally harmful (BROOM, 1996, KEELING et al., 2011).The range of welfare extends from very good to very poor. BROOM (1996) points out that welfare can be reduced without actual suffering. If there is suffering involved the animal's welfare is more severely reduced. However, the extent of how much the animal is suffering can often not be measured (BROOM, 1996). Examples for indicators for good welfare are "normal growth and reproduction" and "no injury" whereas bad welfare can be seen - for instance - when the growth or reproduction rate as well as the life expectancy are reduced. KEELING et al. (2011) agree that impaired reproduction is a sign for bad welfare; however they argue that a high reproduction rate or success do not automatically stand for a high welfare state. On the contrary, these aspects may go hand in hand with increased health problems and reduced longevity (KEELING et al., 2011, see also EDWARDS, 2007).

Another concept of animal welfare deals with the feelings of animals. The main argument is the fact that it is only of importance what the animal actually feels. Welfare therefore is about "conscious states, that is, with the absence of suffering and (probably) with the presence of states of pleasure" (DUNCAN, 1996, see also DUNCAN and PETHERICK, 1991). This concept stands against the presumption that an animal's welfare is about "the satisfaction of needs" but with the emotions that come along with those needs. DUNCAN (1996) argues that a stressed animal may feel stressed in a welfare reducing way but there is also the possibility

that it feels excitement even when the physiological parameters are the same. The associated feelings however may be completely different which is why Duncan focuses on assessing feelings when carrying out welfare assessment (DUNCAN and PETHERICK, 1991, DUNCAN, 1996). However, so far the assessment of feelings has been beyond scientific reach which is one major critical point within this concept. Only in the last years scientific methods have been developed which reveal more valid results and allow clearer statements (e.g. pain assessment). Furthermore, experts point out that positive feelings in animals should be examined and in further consequence, welfare should be determined by these emotions (KEELING et al., 2011).

The third concept relates the welfare of an animal to its natural state, i.e. to what extent an animal is able to show its natural behaviour repertoire. According to this concept, a physiologically healthy animal might have poor welfare if it is restricted in expressing its natural behaviour (FRASER, 2008). LUND (2002) proposed the term “natural living”: this includes feed that is appropriate for a specific animal or species. Furthermore, an environment which is similar to the natural habitat should ensure a high state of welfare (LUND, 2002: 41). In this aspect LUND (2002: 42f) argues that a natural environment most likely involves stressful situations such as the potential appearance of predators in an outdoor system – however natural living is misinterpreted when a farmer stops protecting his or her animals. On the contrary, the farmer has the obligation to protect the animals from any predators. Further animal welfare issues can be found in outdoor systems. Whereas the expression of natural behaviour can be more easily shown in these systems, the animals are more likely to suffer from parasites and weather conditions than in artificial environments (FRASER, 2008). Still, the welfare of an animal which is given the opportunity to express natural behaviour might be effectively enhanced as performing this behaviour is likely to go hand in hand with positive emotions (KEELING et al., 2011).

An overview as well as a discussion about the three different approaches can be found in FRASER (2008). He also refers to a model of animal welfare considering the three above mentioned concepts as complementary approaches (Figure 2).

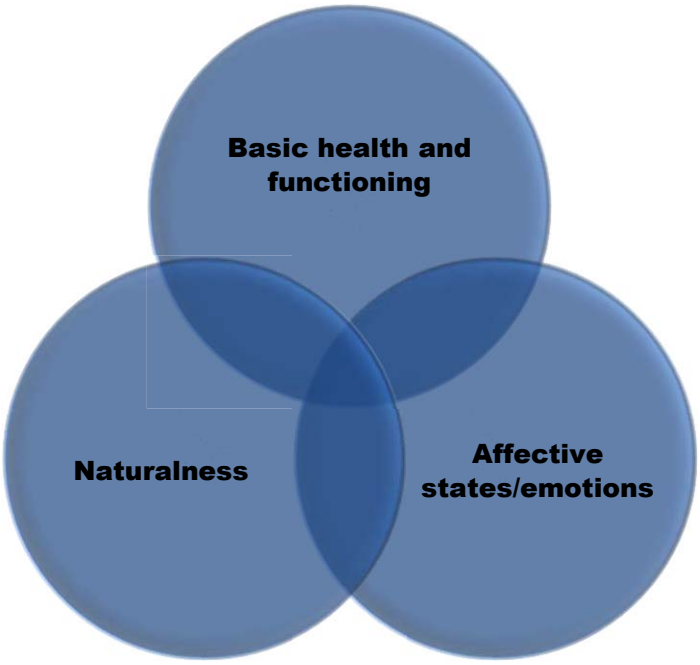


Figure 2: Three concepts of Animal Welfare, adapted from Fraser 2008 and Lund 2002

One of the most important, practical guidelines for ensuring animal welfare was developed by the *Farm Animal Welfare Council* (FAWC). The so-called *Five Freedoms* are currently referred to by several animal welfare organizations such as the *RSPCA* (Royal Society for the Protection and Care of Animals) or *The Brooke*:

> *"**1. Freedom from Hunger and Thirst** - by ready access to fresh water and a diet to maintain full health and vigour.*
>
> ***2. Freedom from Discomfort** - by providing an appropriate environment including shelter and a comfortable resting area.*
>
> ***3. Freedom from Pain, Injury or Disease** - by prevention or rapid diagnosis and treatment.*
>
> ***4. Freedom to Express Normal Behaviour** - by providing sufficient space, proper facilities and company of the animal's own kind.*
>
> ***5. Freedom from Fear and Distress** - by ensuring conditions and treatment which avoid mental suffering"* (FARM ANIMAL WELFARE COUNCIL, 2013).

3.2 Animal welfare assessment

The following chapter deals with animal welfare assessment in general and gives an overview on welfare parameters and principles as well as an insight into welfare assessment in the field.

3.2.1 Welfare parameters

Welfare parameters can generally be divided into two types. The first category is environment- and management- based (in other words input- or resource-based) and deals with aspects such as feeding facilities or space allowance. Assessing this kind of parameters has been described as “fairly uncomplicated” as data are generally regarded to be easily and quickly collected (JOHNSEN et al., 2001). Furthermore, repeatability of resource-based parameters (e.g. length of feeding trough) is usually not considered a problem. Furthermore this kind of data provides a good basis for decisions, which changes to implement in an existing environment (JOHNSEN et al., 2001).

In the past the main focus of welfare assessment of farm animals lay on the resource-based indicators, however in the last 20 years animal welfare assessment developed from a resource-based to a more animal-based approach (WHAY et al., 2003, KEELING, 2005). Animal-based indicators deal with the animal itself, i.e. how the animal reacts to a specific environment. Within this category, animals are assessed regarding their behaviour, health and physiology (JOHNSEN et al., 2001). Nevertheless, the assessment of this kind of parameters can be more difficult compared to the resource-based indicators as firstly, the data collection is usually more time- and resource- consuming and requires more knowledge (e.g. the assessment of species-specific behaviour). Secondly, results obtained by means of animal-based indicators are considered to be less clearly and concretely interpretable and therefore less suitable for welfare assessment which - according to JOHNSEN et al. (2001) – is the case with physiological and behavioural parameters. Physiological measurements such as increased heart rate, adrenal activity or higher level of the glucocorticoids can indicate a lower welfare status however increased levels may also occur due to courtship which does not necessarily comprise low welfare. Therefore, the interpretation of those welfare measures is not so straightforward and must be carried out carefully (BROOM and FRASER, 2007: 58f).

There is a general understanding that resource-based parameters may be more objectively measured than animal-based ones (e.g. WHAY et al., 2003). However this assumption must be questioned: MULLAN et al. (2011) argue that many resource-based measures are assessed in a subjective way and by means of a subjective judgement. The authors give as an example that the assessment of wound length is more accurate and objective than the seemingly objective recording of the sharpness of a protrusion.

Generally the combination of both types of parameters is considered as most useful for a valid welfare assessment. For instance the welfare assessment protocol developed within the WQ© project combines resource- and animal based indicators (e.g. FORKMAN and KEELING, 2009, WELFARE QUALITY® CONSORTIUM, 2009a)

3.2.2 Principles of welfare assessment

Generally, validity and reliability are desirable aspects when assessing an animal's welfare. KNIERIM and WINCKLER (2009) add feasibility as third property.

Validity describes whether a welfare parameter measures what it has been designed for; i.e. to what extent a certain measure actually provides information on the welfare state. Several types of validity may be distinguished. When referring to face validity, experts agree on one measurement of welfare and therefore grant it to be valid (e.g. injuries). For instance, the Welfare Quality® protocol consists of parameters which experts judged to be valid; therefore face validity is assigned to all parameters. Content validity means that all aspects which define a specific welfare issue are included in the measurement. Additionally, a newly developed welfare measure has criterion validity if it has a relationship to an already existing measure whereas construct validity exists when the relationship between welfare and a variable has been proven (SCOTT et al., 2001).

The second property of an assessment tool is reliability which comprises inter-observer, intra-observer and test-retest reliability. Usually, reliability – as the ratio of the variability of one's individual score compared to the variability of the score of all observers - is shown as a dimensionless coefficient and is ranging from 0 to 1. Inter-observer reliability deals with how consistently several individual assessors rate the same animal whereas the intra-observer reliability is looking how consistent ratings of an individual observer are on different occasions. Test-retest reliability focuses on how consistent the ratings are when measuring the same animal on two or more repeated assessments (SCOTT et al., 2001, KNIERIM and WINCKLER, 2009).

3.2.3 Examples for applied animal welfare assessment

The welfare of farm animals was the main focus when developing the *Welfare Quality*® protocols. Four main principles – each defined by several criteria in detail – are addressed in these protocols which are currently dealing with cattle, pigs and poultry welfare. The first welfare principle *Good feeding* includes *Absence of prolonged hunger* and *Absence of prolonged thirst*. Secondly, Good housing covers Comfort around resting, Thermal comfort and Ease of Movement. Good health is the third principle and consists of the following criteria: Absence of injuries, Absence of disease and Absence of pain induced by management procedures. Lastly, the fourth principle – Appropriate behaviour – is dealing with Expression of social behaviours, Expression of other Behaviours, Good human-animal relationship and positive emotional state (WELFARE QUALITY® CONSORTIUM, 2009b: 22).

Another example for applied animal welfare assessment with a focus on working equines has been developed by the *University of Bristol* in collaboration with the equine charity *The Brooke*. A working equine welfare assessment protocol (WEWA) is used in order to monitor the charity's work. Animal based indicators were used in the protocol as in case of working animals – contradicting JOHNSEN et al. (2001) – these indicators are easier to record than the above mentioned resource-based indicators which are mainly dealing with housing conditions etc. (PRITCHARD et al., 2005). However, in the field of equine charities it was common to check the input based parameters such as number of treatments or delivered vaccines rather than to look on the outcomes of these actions taken (LEEB et al., 2003). After the actual assessment experts ranked the detected welfare issues and additionally, risk factors related to these welfare issues were identified. This approach should ensure that the specific welfare issues regarding different species and various work types are targeted in a following intervention carried out by *The Brooke*. The positive and negative outcomes of that intervention would - in further consequence - be recorded with the same protocol (PRITCHARD et al., 2005). This working equine welfare assessment protocol also refers to the requirements for welfare indicators that have been stated by SØRENSEN et al. (2001) with some minor additions in order to be adaptable to the environment of working equines. Welfare indicators therefore "*will*

1. *Describe relevant and significant aspects of what matters from the point of view from the animals*
2. *Express changes over time*
3. *Be capable of being influenced by decisions and actions taken by the individual farmer* (in PRITCHARD et al. (2005) the equine owner or user)
4. *Be measurable in a relatively cheap and easy manner"* (SØRENSEN et al., 2001)

The circumstances within the protocol is applied require that the assessment can be carried out quickly (less than 10 minutes) and without any disturbance during a working day. Additionally, the assessor does not have to touch the animal with most measures (PRITCHARD et al., 2005). The WEWA represents an example for a feasible approach of working equine assessment.

3.2.4 Qualitative Behaviour Assessment

One part of the mentioned Welfare Quality® protocols includes a qualitative approach to assess animals' behaviour. In the recent 15 years this so-called Qualitative Behaviour Assessment (QBA) has been developed and implemented. Within this approach the outcome is not what an animal does but in which matter an animal behaves. Subjective descriptions, which had been reluctantly used by experts beforehand, are part of QBA: an animal can now

be described as content, frustrated, sociable etc. In earlier years such terms were seen as anthropomorphic attribution – i.e. humans impute human behavioural expressions on to animals (WEMELSFELDER et al., 2001). In order to prevent anthropomorphism a certain degree of expertise, skills and knowledge in species-specific behaviour is necessary. WEMELSFELDER (2007: 28) points out that "extensive experience is needed to correctly discriminate categories of behaviour and measure these categories reliably". Expertise and experience are – according to the author - furthermore of importance as the expressive quality of behaviour is easier to detect in some species compared to others.

Originally QBA uses a Free Choice Profiling approach which includes observers using their own set of descriptors or terms and rating these terms along a Visual Analogue Scale. In a further stage these subjective ratings can be statistically analysed and can therefore be compared (WEMELSFELDER et al., 2001, MINERO et al., 2009). However, since the analysis of FCP requires a larger set of observers (i.e. ~ 10) which is not a feasible approach in on-farm assessment the methodology of QBA was further developed. Instead of using own descriptors the observers can be provided with a so-called fixed terms list. For instance, in WQ© 20 descriptive terms are used for the qualitative assessment of cattle, pig and poultry behaviour (WEMELSFELDER et al., 2009a, WEMELSFELDER et al., 2009b).

3.3 Donkeys in developing countries

The following chapter gives an overview of the donkeys' use, importance and welfare in developing countries focussing on the situation in Ethiopia. This chapter also summarises experiences and results from published donkey welfare assessment studies.

The world donkey population is around 42 million animals. China has the largest population followed by Ethiopia and Pakistan (FAOSTAT, 2013), which is summarized in Figure 3.

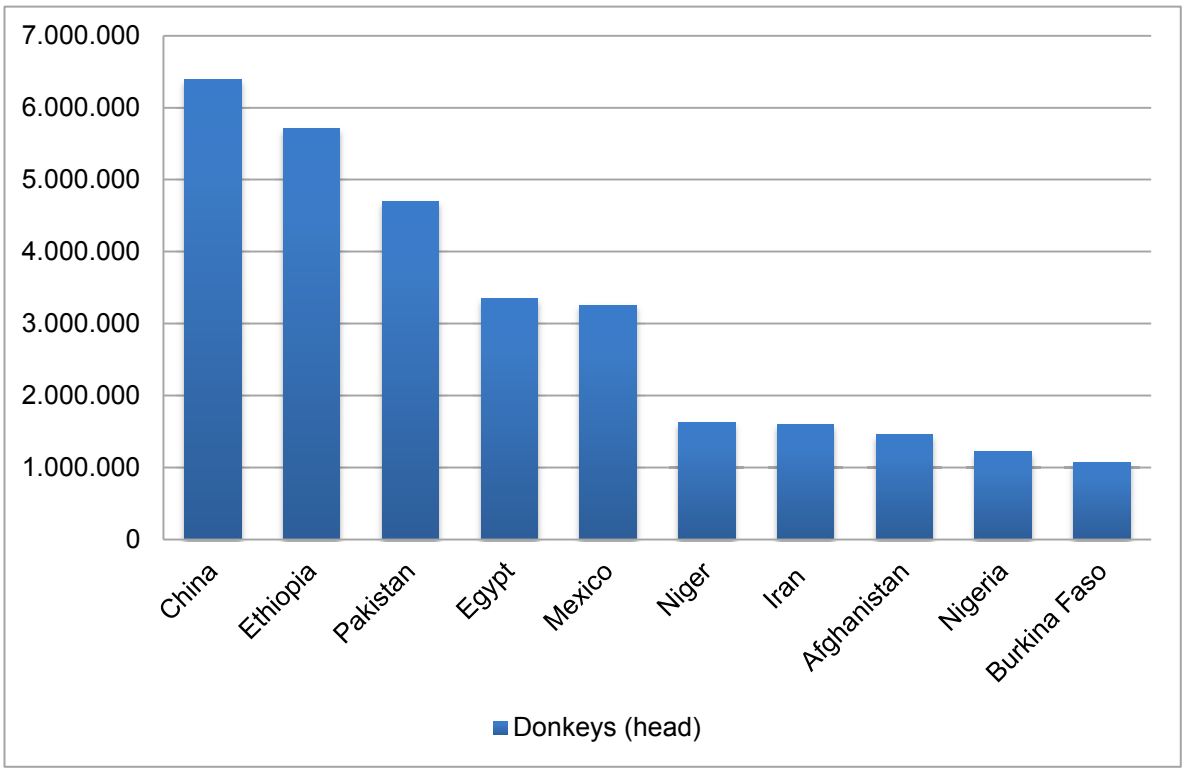


Figure 3: Largest donkey populations in 2011 (FAOSTAT, 2013)

Interestingly, information on the donkey's domestication and distribution is quite scarce despite the fact these animals are wide spread in Africa. Furthermore it is not fully resolved which wild ass (i.e. Nubian wild ass, African wild ass) is the ancestor of the modern donkey. The reason for the domestication of the wild donkey is still unclear: it cannot be assumed that it was only used for transport as it is today – the consumption of its meat and milk might also have played a role (BLENCH, 2000). BLENCH (2004) speaks of a contradiction as the donkey is considered as a highly important livestock whereas there are hardly archaeological remains such as rock art representations or historical-textual documents. He sees the ever existing low status of these animals resulting from the donkey's use of poor small scale farmers as the probable reason for the lack of information regarding donkeys. The ancient Egypt forms an exception: firstly, there are sources such as wall paintings that show donkeys used in agriculture (BLENCH, 2000) and secondly, recent archaeological findings point out that these animals were regarded as highly valuable transport animals and were not kept for meat consumption (ROSSEL et al., 2008).

Regarding domestication a further aspect becomes relevant regarding animal welfare. Equally to other domesticated animals (e.g. STUDNITZ et al., 2007)donkeys' behaviour and needs are the same as those from their wild ancestors. Both the wild ancestor as the domesticated donkey would run away from danger – only if there is no other possibility they start to kick and bite. Additionally donkeys are more subtle in their fear-related behaviour:

whereas horses are openly fearful in novel situations a donkey's reaction is often mistaken for stubbornness rather than fear. Another relevant aspect influencing the welfare of donkeys is the strong bond that these animals develop to each other. Donkeys prefer to be kept with a companion to whom they form lifelong friendships. As their bond tends to be very strong separating befriended animals can lead to stressful situations and distress (THE DONKEY SANCTUARY, not stated-b: 6, 8, JONES, 2012 personal communication). MURRAY et al. (2013) refer to anecdotes of donkey owners and experts and tested pair bonding and companion recognition with 55 donkeys. They report that donkeys do not choose their companions deliberately but they in fact have non–random nearest neighbour relationships. In the Y-maze test donkeys prefer approaching and standing close to their companion compared to a familiar and unfamiliar donkey. The authors also conclude that these results support these anecdotal references that separating befriended donkeys can cause stress and should be prevented (MURRAY et al., 2013).

3.3.1 Donkey use and role in developing countries

In Africa donkeys are mainly kept for work, mostly as pack but also as draft animals which is highlighted by the common expression as "beast of burden" (PEARSON et al., 1999). They are also part of ceremonies such as weddings but also play a role in tourism (PRITCHARD et al., 2005, BLENCH, 2000). To a lesser extent and depending on the region donkeys are also kept as breeding animals (i.e. solely kept as breeding donkeys with the aim to produce offspring), for milk production and meat consumption. However, meat consumption is according to BLENCH (2000) highly underestimated as eating donkeys is very often a taboo for instance in the Islamic religion and certain Christian groups.

The role of the donkey in developing countries cannot be underestimated although their economic value is difficult to assess. Nevertheless, donkeys are described by experts to be economically beneficial although hard figures are still missing (FAO/THE BROOKE, 2011: 7). One important aspect why the donkeys' socioeconomic value is mostly underrated derives from the existing paradigm of "development" which includes industrialisation, motorisation and modernisation. The "work force donkey" does not fit into this paradigm. A large proportion of people in the developing world is not part in the so called "modernisation process". Many of these people work in small scale farming and transport donkeys are still a highly important work force. For example, in Ethiopia 85 % of the population are subsistence farmers and are heavily dependent on the donkey as a transport animal as three quarters of these farmers are living more than six hour walk away from the nearest road. It also represents the main agricultural work force in Ethiopia (POWELL, 2004, TESFAYE and MARTIN CURRAN, 2005).Furthermore these animals play a vital role for women as donkeys relieve the "everyday work" such as fetching water but can also contribute to generating additional

income (FERNANDO and STARKEY, 2004). In this context the lack of regard for and knowledge about these animals shown by authorities, governments and research confirms the overall low image the donkey has in the eyes of these stakeholders (FERNANDO and STARKEY, 2004, FAO/THE BROOKE, 2011: 8, 15, PEARSON et al., 1999, in this context see also BRYCESON, 2010).

Despite the important role of the donkey its reputation and status in Africa is poor partly resulting from their sole output work. This low image is related to the above mentioned neglect from authorities and research (PEARSON et al., 1999). Traditional sayings also give a hint about the image of the donkey. BEKELE (2004) collated 13 Ethiopian sayings, consisting of nine negative and four positive ones. The negative ones mainly describe the donkey as weak, fearful and careless as well as that their meat is not for consumption. Contrariwise, the positive sayings put the focus on the donkeys' strength to ensure a family's daily income and living (BEKELE, 2004). Another quite often quoted saying with various modifications is "A man without a donkey, IS a donkey" (THE DONKEY SANCTUARY, 2013) or "Without a donkey, my wife and I become the donkeys" (POWELL, 2004) pointing out the importance of working donkeys. Despite its obvious importance the donkey's low image seems to be cultivated by non-donkey owners (farmers and officials) and therefore, any attempts to improve the status and welfare of the donkeys is not encouraged (PEARSON et al., 1999).

3.3.2 Donkey welfare

Due to the topic and focus of the present thesis, the literature presented on donkey welfare mainly deals with donkey welfare in developing countries summarising studies on donkeys as well as studies on "working equines", which usually include horses, donkeys and mules. There are various animal health and welfare issues that are often specific to the regions and work type of the animals (BURN et al., 2010a, PRITCHARD et al., 2005, MEKURIA and ABEBE, 2010).

The following parts follow the order of the Hands-on Donkey welfare assessment protocol describing each finger in detail.

Behaviour/Demeanour

According to BURN et al. (2010b) several behavioural indicators shown by working equines could be useful indicators for – underlying – health and welfare problems. It may also provide some information on the quality of the human-animal relationship.

Regarding unresponsiveness of working equines it is not considered as a state of relaxation and contentment but of total exhaustion and having reached the individual's limits or exceeded those limits (see also POPESCU and DIUGAN, 2013). An apathetic or severely

depressed state of working equines is a problem in many regions. In India 21 % of 110 examined horses showed this kind of behaviour (BROSTER et al., 2009). 11.5 % of 2596 donkeys examined across developing countries were apathetic or severely depressed (PRITCHARD et al., 2005) whereas 13.1 % (n=5481) donkeys expressed signs of apathy and depression in another study carried out in nine developing countries. How donkeys react to observers is also of interest. In general, the human-animal relationship influences the well-being of animals. Limitations in productivity and welfare are caused by fear of humans (HEMSWORTH, 2003). Furthermore fearful animals are less easy to work with. 12.1 % of 5481 donkeys (PRITCHARD et al., 2005) were friendly towards an unknown observer whereas the rest was either aggressive (44.3 %) or was not responding at all (43.6 %). Around 90 % were responding when the observer walked down their side. A behaviour only shown by donkeys is the so-called tail tucking (clamping down the tail or tucking in the hindquarters according to PRITCHARD et al. (2005) which is said to be an indicator for having experienced being beaten; 28.2 % had their tails tucked in. Additionally, the acceptance of chin contact has been tested with 18.3 % of the donkeys avoiding chin contact.

Investigating working horses in Romania (n=715), POPESCU and DIUGAN (2013) extended the categories of reactions towards humans into aggressiveness, avoidance/fear, indifference and friendliness. Additionally, they tested the behaviour towards humans with an unknown person as well as the handler based on the assumption that the animals react differently to known and unknown humans. The authors found significant differences of the animals' reaction towards the assessor and the owner. Indifference was most prevalent during the observer approach for both known (74.3 %) and unknown handlers (52.3 %). The authors considered the percentages of fearful/avoiding (20.6 % and 11.1 %, respectively) or aggressive animals (2.9 % and 2.4 %, respectively) as high as they argue that the approach test should be seen as a neutral stimulus under these working conditions. During the Walk Beside Test – considered as another neutral stimulus according to the authors - fear/avoidance towards observer and handler was most prevalent, although with a significant difference between observer (63.6 %) and the handler (48.5 %). The authors considered this as highly important since this behaviour might be based on negative experiences and could derive from improper handling. Additionally, this behaviour is considered to go often hand in hand with aggressiveness which was relatively often displayed in the chin contact test during this study: almost 8 % showed aggressive behaviour towards both the owner and the assessor (POPESCU and DIUGAN, 2013).

BURN et al. (2010b) report a correlation between the diverse parameters testing aversion against humans. Donkeys with their tails tucked in significantly avoided the observer approaching or displayed aggression and touching the chin. The authors point out that these tests do not suggest a bad human-animal relationship (i.e. fear towards humans) as there

were no correlation found between behavioural tests and man-made problems such as wounds on the hindquarters due to beating or slit nostrils. Additionally, limitations are given due to the unfamiliar observer: displayed behaviour might be different if a familiar person is approaching the animal. As the behaviour pattern is consistent in each individual animal (e.g. avoidance behaviour throughout all tests) and is considered fairly prevalent within the study of BURN et al. (2010b) the authors conclude that more research specifically dealing with the question if animals show fear of humans or of novel stimuli is needed. POPESCU and DIUGAN (2013) report a difference of behavioural reaction towards unknown and known observer in working horses, however there are no studies dealing with working donkeys so far.

There is a correlation between various behaviour and health parameters. For instance, lack of unresponsiveness towards environment/handling was weakly but significantly correlated to low BCS, abnormal colour of mucous membrane, lesions of skin and deeper tissues and abnormal gait (PRITCHARD et al., 2005, BURN et al., 2010b). As mentioned above, apathy is considered to relate with various health and welfare issues such as low BCS and probably a considerable number of severe skin lesions.

There have not been any (published) studies done yet regarding the qualitative behavioural assessment in donkeys. However this method was used to assess the welfare of horses and ponies kept as companion animals (NAPOLITANO et al., 2008). Two main dimensions of expressive behaviour – namely “quiet/nervous” and “attentive/bored” - were described. Interestingly horses were perceived as more quiet but also more curious and attentive than ponies. MINERO et al. (2009) tested the relationship between quantitative and qualitative parameters by assessing foals’ responses to an unfamiliar handler. The authors argue that qualitative assessment gives additional information to merely quantitatively collated data by giving an interpretation of the latter ones (see also FLEMING et al., 2013).

Body Condition Score (BCS)

Many working equines are very thin which is considered as a major welfare issue, also within the context of the working conditions. The BCS is considered as good indicator for several welfare issues: skin lesions, gait and sole abnormalities and faecal soiling are more likely to be found in/associated with thin animals (BURN et al., 2010a). Also a correlation between a low BCS and parasite infestation (AYELE et al., 2006), as well as malnutrition and overwork (BURN et al., 2010b) has been described.

Around 70 % of 5481 donkeys in five developing countries are reported to have a BCS of 1 or 2 (1 = very thin, 5 = very fat) (PRITCHARD et al., 2005). A study carried out in Ethiopia and using a six point scale (0= very thin, 1 = thin, 2 = fair, 3 = medium, 4 = fat, 5 = very fat) showed that more than 25 % had a BCS of 0, 1 or 2 although there is significant difference between thin horses (59.8 %) and thin donkeys (16.3 %) (MEKURIA and ABEBE, 2010). 26 %

of donkeys presented for veterinary treatment in Mexico were assessed as poorly conditioned (BCS ≤ 2) (BURDEN et al., 2010).

PRITCHARD et al. (2005) report a correlation between low BCS and systemic health abnormalities, abnormal gait and limb problems. A correlation between lower BCS and dehydration was found by MCLEAN et al. (2012)

Lameness

Lameness is considered to be a widespread problem across several countries and work types and is considered to involve pain (ROSS, 2011, BROSTER et al., 2009). This is insofar a highly relevant welfare issue as working equines are working for several hours per day and therefore are likely to suffer from pain throughout this time (PRITCHARD et al., 2005). BROSTER et al. (2009) also point out the correlation between lameness and low body condition. Citing several studies the authors are discussing the causal direction and conclude that lameness can be caused by a low body condition but also vice versa. According to a study based on data from 9 developing countries, gait abnormalities were highly prevalent with nearly 96 % of the working equines showing signs of lameness (BURN et al., 2010a). Also PRITCHARD et al. (2005) found nearly 95 % of the examined donkeys (n = 2596) to show a gait abnormality and the authors point out the high prevalence of limb abnormalities. Similar findings can be seen in BROSTER et al. (2009) who examined 227 horses in Pakistan and India and found all animals to be lame. Additionally, 87 % of the examined animals had a leg score of 3 or 4 on at least one leg (0 = sound, 4 = non weight bearing).

The causes of lameness are differing depending on the region. MORGAN (2006) shows that traumata caused by road traffic accidents are the main reason for lameness in urban areas whereas in rural regions lameness occurs because of wounds (mainly hyena and donkey bites).

Wounds

Wounds are reported as a serious welfare issue in working equine welfare populations. Improper harness is considered to be a major cause for lesions and wounds (BURN et al., 2010a, BURN et al., 2008). Furthermore, the work type plays a role as does the handling of the animals: for instance, draft animals usually have more lesions at the commissures of lips than pack animals as there are no bits used with the latter. However, pack donkeys working in brick and construction sites more likely have wounds at the hindquarters and tail base as they tend to get beaten in order to be moved (PRITCHARD et al., 2005, see also SELLS et al., 2010). Another cause for wounds on the legs is the so-called hobbling when two limbs are tied together in order to prevent the donkey from moving away. Tethering and hobbling wounds are described as highly prevalent (BROSTER et al., 2009). Problems for the welfare of

animals affected with wounds arise from pain and/or infection. BURN et al. (2010b) consider wounds associated with apathy: the affected animals are apathetic because of pain and/or infections and vice versa. Apathy may also lead to wounds as the animals are more likely to have traffic accidents.

TESFAYE and MARTIN CURRAN (2005) reported an increase of back sores during the rainy season (May till October) in Ethiopia and assumed a connection between wet circumstances and reduced wound healing. They found 58 % of 33 donkeys to have back sores.

Wounds are also considered to correlate with the dehydration status (PRITCHARD et al., 2005, see also chapter 3.3.3). Another correlation shown in several studies exists between low BCS and wounds as the natural padding due to a loss of muscles and fat deposits likely leads to lesions and wounds caused by harnesses (PRITCHARD et al., 2005). The donkeys observed (n=2596) in the same study had skin lesions and/or lesions of deeper tissues at the withers (10.2 %) and hindquarters (12.1 %). Also the girth area was affected with 18.3 % of the donkeys having a lesion there. SELLS et al. (2010) report an overall prevalence of 54 % of 147 observed pack animals describing the withers as the most prevalent site. MEKURIA and ABEBE (2010) found tether and hobbling lesions to be highly prevalent in working donkeys in Ethiopia: 93.4 % of 381 animals were affected by this kind of lesions. Lesions of the skin and/or deeper tissues on other body parts were not that highly prevalent showing 12.9 % and 16.0 % of the donkeys having a lesion on the wither/spine and tail/tail base, accordingly. Injuries (defined as grossly visible skin/tissue damages) were found in 79.4 % of the examined donkeys (n=320). The majority of those injured donkeys had severe injuries (BIFFA and WOLDEMESKEL, 2006).

Other Signs of Disease or Injury

Research on parasites and parasite infestation in working equines in Africa has been carried out quite intensively compared to studies on health and welfare. Working equines are suffering most likely from parasites; strongyle infestations are very common (e.g. GETACHEW et al., 2008, AYELE et al., 2006). An abnormal coat condition is considered as another sign of impaired welfare whereas a healthy animal usually has an even and flat coat (THE DONKEY SANCTUARY, not stated-b: 16). Additionally, the coat condition can be an indicator for parasite infestation (THE DONKEY SANCTUARY, not stated-b: 20). PRITCHARD et al. (2005) report that 2.5 % of the 2596 donkeys had a staring/matted/dry/uneven coat.

Eye problems are also highly prevalent according to PRITCHARD et al. (2005) as 64.4 % of observed donkeys (n=2596) showed an eye abnormality. However, other studies have not discovered such a high prevalence (BURDEN et al., 2010, SCANTLEBURY et al., 2013).

Other diseases such as Rabies and Tetanus do play an important role in developing countries (KAY and KNOTTENBELT, 2007)

3.3.3 Additional aspects of donkey welfare assessment

Welfare assessment of working equines differs in some aspects from farm animal welfare assessment but also horses, which are kept as companion or sport animals. For instance any resource-based assessments dealing with the housing of pack donkeys on markets are difficult to obtain as these assessments are usually carried out on their "working environment". Furthermore health records for inspection are not available for these working animals. An additional example is the lameness scoring which is usually carried out by letting the horse trotting 20 m away and back to the observer. This is according to PRITCHARD et al. (2005) not possible in a working equine environment due to limited time for the assessment in the field and the crowding in observation areas. Furthermore the animals are not trained to trot in-hand which is common in companion animals. Therefore, PRITCHARD et al. (2005) assessed a gait abnormality by observing the animals for 12 paces at walk.

In order to get a reliable statement about observer reliability a balanced sample is needed, which should include a wide range of scores to be able to conclude if different individuals rate in the same manner. In an unbalanced population – e.g. with a prevalence of 95 % for lameness and a high inter-observer reliability – it is not clear if the high percentage derives from a general good agreement or by chance as it is hardly possible to rate differently (BURN et al., 2009). As a matter of fact, working equine environments generally show very high prevalence for various welfare problems as lameness and low BCS (PRITCHARD et al., 2005, BROSTER et al., 2009). In a study carried out by BURN et al. (2009) the observer reliability for working equine welfare assessment was tested by comparing two groups of assessors in India and Egypt. The authors state that the overall reliability for donkeys is significantly lower than for horses however they do not provide any detailed explanations for these differences. Variables such as horn quality, mucous membranes and general attitude showed low percentage of agreement for donkeys. Behavioural assessments display various inter- and also intra- observer reliability ratings. This could also be due to a change of behaviour between different observations. Interestingly, the variable gait (0= normal, 1= abnormal gait) had a low agreement in horses; therefore BURN et al. (2009) suggest a more differentiated lameness score in future studies in order to get more information about the lameness situation in working equines. The authors conclude that an appropriate welfare assessment is needed due to high prevalence of many working equine welfare issues in order to develop intervention and improvement strategies. In this aspect they also suggest to artificially pre-select a sample which includes the whole range of degrees of different parameters.

Another topic of discussion is the validity of parameters as welfare indicators: A good example is the assessment of dehydration status of working animals PRITCHARD et al. (2005) addressed this issue by means of assessing signs of faecal soiling on hind limbs, skin tent

(loss of elasticity) and heat stress (see also PRITCHARD and WHAY, not stated). However, the validity of the skin tent test has not been confirmed and needs further research. PRITCHARD et al. (2008) compared the skin tent test to other dehydration parameters such as drinking behaviour (water intake) and plasma osmolality. The authors found no correlation between these parameters. Instead, the skin tent test was influenced by the side the test is carried out on the animal, the anatomical location as well as the coat moisture (ranging from dry to wet). Furthermore, the skin tent test showed variable agreement between different observers according to PRITCHARD et al. (2007). The percentage agreement ranged from 40 % to 99 %. The authors discuss several reasons for this result: variable assessment by each observer (despite training and guidelines), the individual previous experience, the subjective demarcation between scores and the biological variability. Contrariwise, the heat stress behaviour consisting of increased respiratory rate, increased respiratory depth, head nodding, flared nostrils and apathy correlates with increased body temperature is considered as valid parameter and can therefore be used as a welfare indicator (PRITCHARD et al., 2006, PRITCHARD et al., 2008).

An additional aspect of low welfare is pain and pain assessment. Publications from recent years demonstrated a lack of knowledge about pain behaviour of donkeys (ASHLEY et al., 2005). In order to get more information on that aspect ASHLEY et al. (2006) worked on the development of an ethogram of potential behavioural signs of pain in working donkeys. That ethogram should also help to assess pain in working equines in the future. The authors compared donkeys in the UK with working donkeys in India, Pakistan and Dehli because the behaviour of healthy, non-working animals was assumed to differ from those suffering from multiple and severe health and welfare problems. By observing the animals the authors found significant differences in ear position, head position, weight-shifting and tail swishing. Additionally, they assumed that the absence of play behaviour shown by the Indian and Pakistani donkeys might be an indicator for pain among other factors such as exhaustion. These contributing aspects as well as the stoic nature of donkeys make it according to the authors difficult to describe definite behavioural pain indicators (ASHLEY et al., 2006). ROY et al. (2010) identified the following pain behaviours which are - according to the authors - comparable to the behaviour of horses being in pain: working donkeys were given a Non-Steroidal Anti-Inflammatory Drug (NSAID) respectively a placebo in order to compare the behaviour before and after the administration. The results show a significant difference between the two groups: the NSAID group was performing more walking and chewing/biting bouts as well as more sniffing. The authors therefore conclude that the decrease of exploratory behaviour is a potential indicator for assessing (the absence of) pain behaviour (ROY et al., 2010). Another approach to gain more information about the pain-related behaviour of donkeys was developed by OLMOS et al. (2010): the researchers compared pre-

mortem behaviour and clinical conditions of donkeys that were destined for euthanasia with post-mortem pathological findings. They formed two lists one containing clinical pain indicators also including behavioural signs and the other consisting of pathological findings which are said to be related to pain. The aim is to form associations between the different fields such as correlations between ante-mortem behaviour and post-mortem pathological findings such as trauma, acute inflammation or swelling within a confined area (e.g. cerebral swelling) (OLMOS et al., 2010).

4 Animals, study design and methods

4.1 Adaptation of the Hands-on Donkey Welfare Assessment Protocol

The development and testing of the Hands-on Donkey welfare assessment protocol was carried out in collaboration with The DS in UK and Ethiopia.

The DS was founded by Elisabeth Svendsen and is operating in UK, Ireland, other European countries and as well as overseas. In Ethiopia a DS branch was established in co-operation with the University of Addis Abeba and is now running several projects throughout the country. The main aims are to provide veterinary treatment for donkeys and mules as well as training (e.g. harness making). In Debre Zeit the office and the clinic are located on the compound of the university (THE DONKEY SANCTUARY, not stated-c).

The Hands-on Donkey Welfare assessment protocol, which served as starting point for this study, is based on the experiences of the DS's work in their operating areas. The idea of using a hand as an illustration was developed by Stephen Blakeway (Director of International Relations) and his team as he used to count the main donkey welfare issues on the fingers of one hand. The fingers resemble the major problems which relate to behaviour/demeanour, body condition, lameness, wounds and other signs of injury and disease. It has been used in a qualitative, subjective way which included an individual, subjective estimation of the worst and best place to live for a donkey (1=worst place to 5=best place) with regard to the five fingers. No defined, detailed welfare parameters were used for this approach.

In order to do a more detailed assessment the next step was to identify parameters to the respective finger. For that purpose existing equine welfare assessment systems as well as scientific and non-scientific literature were researched. The following references were then considered:

- DEWHURST and HENSTRIDGE, not stated, CANNON and PERROTT, not stated
- PRITCHARD et al., 2005, PRITCHARD and WHAY, not stated
- POPESCU and DIUGAN, 2013, MCLEAN et al., 2012

A first draft of a modified protocol together with a guideline for the Hands-on Donkey Welfare assessment protocol was developed before the field research in Ethiopia and then discussed together with the local staff.

The final updated version of the protocol was finalised in Ethiopia. Based on the experiences of the local Ethiopian DS staff changes were made to detect/describe welfare problems in more detail. Additionally, a guideline was developed to ensure a standardised way of assessment. The protocol was created in Microsoft Word and expands on two pages. The

DS team proposed to use one sheet per donkey in order to prevent missing or made-up values. As observers would use one sheet per donkey they could - according to the DS team - more easily check if they have forgotten to assess one specific part as this would be clearer and visible with one quick look. Furthermore the DS team argued that observers would not easily fabricate data. According to the team's experience assessment sheets with columns for several donkeys are not useful as observers more likely tend to made up field work and pretend to have assessed donkeys.

Besides the development and actual welfare assessment there were two other goals of this project. After having been trained to use the Hands-on Donkey Welfare assessment protocol two different observers (all members of the DS team in Debre Zeit as well as the author) were tested on the consistency of their ratings of 30 donkeys (IOR). Furthermore, QBA was also carried out with the DS team of Debre Zeit. As QBA has never been carried out with donkeys before, interesting additional information was expected to be included in the final protocol.

At the beginning of the field work (i.e. donkey welfare assessment and IOR) the protocol was pilot tested and discussed and several parameters had to be changed since they were regarded inappropriate in the field. For instance, several lameness parameters were concluded into one. In the first drafts of the protocol lameness parameters included "toe in", "toe out", "hoof overgrown" and "hoof too short" as well as "whole limb deformities". As hoof problems were given clearly more emphasis than limb deformities due to the more detailed individual parameters it was argued to conclude these parameters into one, specified as "hoof abnormal". Furthermore the detailed listing of endemic diseases such as rabies or tetanus was withdrawn as proper diagnosis would have required more efforts than possible during a short on-site welfare assessment.

4.2 Inter-Observer Reliability Testing

In order to check the consistency of ratings between different assessors (see 3.3.3) two team members of the DS and the author applied the protocol, that had been finalised after discussions and pilot testing, to 32 donkeys. The IOR testing was carried out in August/September 2013 around Debre Zeit, a town located 50 km southeast of Addis Ababa. Each person scored each donkey one time, i.e. each donkey was assessed in total three times but at different times. Each assessor was asked to do the

Photo 1: Donkeys on market

assessment independently and to not talk to each other during the assessment. The order was not strict – therefore observer 1 did not necessarily start with donkey 1 and continued with 2 but instead chose the next available donkey of the marked donkeys (Photo 1).

The assessments were done on three market sites and mill houses in those locations where donkeys are usually kept between their arrival and departure to/from the market. The animals are usually tied together or tied to a stone either with their leg(s) or their neck(s) (Photo 1). On all three sites, they were not fed or watered during the assessment.

The author chose the sample for the IOR with the purpose to include different welfare and health states of donkeys (i.e. several wounds on back, no wounds on back, hoof problems). Therefore, the sample was not randomized. The donkeys were marked with numbers using a livestock marking stick.

Each of the assessors was handed out a guideline and the protocol form (x 10). As the two DS team members had been accompanying the author during earlier assessments they already had an insight into the procedure. The procedure was talked through again and open questions were answered before the assessment. However in the first round IOR testing one member used half scores with the BCS thus not following the guidelines. Due to the limited number of donkeys assessed the original data was still used. Furthermore there were ambiguous guidelines in the lameness chapter. These ambiguous results could be corrected: the author inquired per e-mail on which basis the two observers of the DS team rated lameness and could therefore assign the correct scores to the respective parameters. Therefore this data could be used for IOR testing.

The data obtained for the IOR was analysed using SPSS. Following BURN et al. (2009) nominal variables with more than two categories (e.g. observer approach) were separated into their binary components (e.g. observer approach “donkey does not move”, observer approach “donkey moves away”). The results are shown in **percentage agreement** (PA) and **Cohen’s Kappa** (k) as well **PABAK**. **Weighted Kappa** was calculated for ordinal variables such as BCS. Due to occurring problems with SPSS when calculating weighted Kappa the tool on the website http://vassarstats.net/kappa.html (last access: 23.04.2014) was used instead and the quadratic weighted k-value is presented as all levels of disagreements were regarded as equal. For instance, with linear weighted k-value the level of disagreement would regarded as less serious when a donkey is scored age 0-5 by observer 1 and age 6-10 by observer 2than when scored age 0-5 by observer 1 and age >15 by observer 2 (SIM and WRIGHT, 2005).

PA was considered sufficient when more than 75 % agreement in nominal variables was achieved. Whereas PA lower than 75 % was still considered sufficient with ordinal variables because “expected percentage agreements decline rapidly as the numbers of possible

scores increases, without necessarily jeopardising clinical relevance"(BURN et al., 2009: 180). Kappa values over 0.4 were regarded as clinically relevant. Furthermore - according to SIM and WRIGHT (2005) - the **prevalence index** (PI) and the **bias index** (BI) along with the Kappa values are presented.

4.3 Application of the Hands-on Donkey Welfare Assessment Protocol

The modified Hands-on Donkey Welfare assessment protocol was applied to 107 donkeys at in total eleven markets and mill houses around Debre Zeit. The Donkeys were commonly tethered by one leg, but it was ensured that any behavioural response as described in the guidelines could be carried out by the animal (Photo 1). For the assessor's safety as well as for practicability reasons a second person held the donkey's head after the behaviour assessment.

Regarding lameness, the assessment included looking for signs of inflammation on the two right limbs and if the sole surface on both right hooves are normal. As this was not feasible in all cases as the animals often did not tolerate to be touched on the hind legs there are two results for these parameters including data from the front legs only and data from both legs.

The data obtained was then entered into an EXCEL spread sheet and the prevalence (in %) was calculated. In order to be able to enter data for the assessment of wounds the author discussed with the Ethiopian DS team on how to define body regions. It was decided to have the following structure: head, neck, back, wound under tail, hind leg, front leg, breast (predisposed spot for wounds), rest of belly and breast. The different wound sizes and depths were entered separately in order to analyse the different wounds and facilitate interpretation.

4.4 Qualitative Behaviour Assessment as an additional approach to donkey welfare assessment

The aim was to carry out a subjective behavioural assessment of working donkeys using a Qualitative Behaviour Assessment (QBA) approach. Within this project this approach included two phases. Firstly – comparable to the Free Choice Profiling approach (WEMELSFELDER et al., 2001)- terms describing the expressive quality of the donkeys' behaviour were generated to create a **fixed terms list** (**Phase 1**). Here, the aim was to consider as many different aspects of the donkeys' behavioural language as well as to practise the personal, subjective judgement. In **Phase 2** the fixed terms list was applied to both a group of donkeys and individual donkeys.

Before carrying out Phase 1, the author explained the method and aim of QBA as it had never been carried out before by the DS team.

Phase 1 took place on several locations such as different markets and watering places and included a five minute observation of a group of donkeys. After these five minutes the participants were asked to turn around and write down terms that would describe what feelings the donkeys express and how they behave.

Places of the QBA phase 1:

22/8/2013: Dukem market (2 groups) – pack donkeys tethered to each other in groups of ~ 5

Photo 2: Donkeys at Dukem market 1

Photo 3: Donkeys at Dukem market 2

27/8/2013: Boset market – tethered pack donkeys, some feeding stuff available

Photo 4: Donkeys at Boset market 1

Photo 5: Donkeys at Boset market 2

16/9/2013: Around Debre Zeit – herd of grazing donkeys

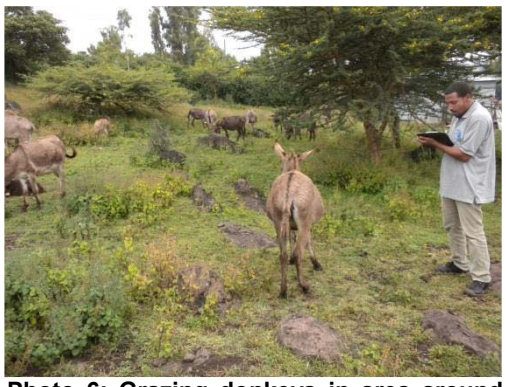


Photo 6: Grazing donkeys in area around Debre Zeit

The team in Phase 1 consisted of three regular DS team members (one time four team members) and the author. One member of the DS team could not speak English and therefore his observations were written in Amharic and then translated by his colleagues.

After having completed Phase 1 the descriptive terms found were collated and those expressions mentioned at least twice by at least two observers were put on a list (i.e. **fixed terms list**) and discussed before carrying out Phase 2. A visual analogue scale was added to each term (Figure 4).

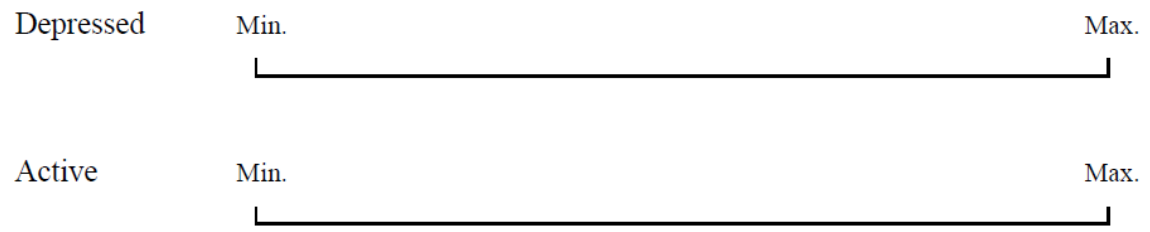


Figure 4: Example for fixed terms list

Phase 2 was carried out twice, on individual animal level (n=12) and group level. The individual animal level was part of the IOR testing as regards the QBA approach as well.

The data derived from Phase 2 of the qualitative assessment was put into an EXCEL file by measuring the distances on the visual analogue scale (starting from min). Observer 2 scored one term twice on one scale. Here, for analysis that value was chosen which in combination with the other values of the given rating corresponds to his usual pattern of scoring. The next step was a Principal Component Analysis (PCA) that was carried out with SPSS. The PCA was calculated for the overall dataset as well as for the individual observers. The aim of these analyses is a reduction of components or dimensions to facilitate interpretation and also to get a better insight into the individual pattern of scoring regarding each observer.

In order to carry out an overall PCA the minimal requirements are the Kaiser-Meyer-Olkin Measure of Sampling Adequacy and the Bartlett's Test of Sphericity. The KMO Measure was 0.675 which was considered as suitable and the Bartlett's Test was significant. Regarding those requirements the collated data proofed to be suitable for the PCA.

For the observer-wise calculation the KMO measures as well as the Bartlett's test were calculated. The KMO Measures were 0.518, 0.560 and 0.600 which was considered as suitable as also the Bartlett's Test was significant.

Furthermore the initial and cumulative eigenvalues were extracted as well as the loadings for each factor on the main components. Only those loadings exceeding $|0.6|$ were considered. Additionally observer-wise individual calculations for each component were carried out and

compared in a table with the calculations of the overall components. One further step was the individual calculation of scores for each observer and each animal.

Furthermore the regression factors for each component were obtained and then displayed in a scatterplot with each donkey pointed out by a number. The different observer scores are shown in different colours.

A last round of QBA was carried out by observing a group of donkeys. The obtained data was not statistically analysed. It was used for a discussion about the experiences with QBA for each individual observer.

5 Results

5.1 Final version of the Hands-on Donkey Welfare Assessment Protocol

Two documents were produced before and during the field research in Ethiopia: firstly a welfare assessment protocol which was based on collated scientific and non-scientific literature. By examining this literature parameters were chosen which were considered to be useful and appropriate to resemble and define the five fingers (behaviour/demeanour, body condition, lameness, wounds, other signs of injury and disease) of the original Hands-on Donkey Welfare assessment protocol. Several versions of a donkey welfare assessment protocol were then produced, discussed with the DS team in Ethiopia and adapted to a final version (see Annex, p.79). Secondly, first drafts of donkey welfare assessment guidelines were developed before the field research in Ethiopia and finalised during the field work. These guidelines were supposed to serve as a handbook which supports observers during donkey welfare assessment if questions arise for instance regarding definitions. After the field work the handbook had to be revised as photos had to be added to the written explanations to make it more concise (e.g. sole surface, coat condition). Furthermore explanations regarding the lameness parameters had to be changed as the definitions were unclear and misleading. This was at least partly due to the underlying concept of the protocol which is to state whether some problem is absent or present. If the parameter is absent and therefore the donkey is assessed as "normal" within this parameter the score is 0. Vice versa if the donkey shows an abnormality the score is 1. This approach was also explained during the training. However – when looking at the protocol form applied in Ethiopia (see Annex) – there is an inconsistency for instance as it says "weight shifting" and "weight bearing"[1]: the first parameter explains an abnormality but the second one a normal condition. When scoring 0 for "weight shifting" that would mean there is no problem. However "weight bearing" is misleading as this describes the normal condition. Therefore scoring 0 could either mean that the observer followed the approach as explained in the training (there is no problem with "weight bearing") or instinctively scores 1 as weight bearing is present. Furthermore the guidelines developed for welfare assessment also include the explanation which ultimately leads to a false – as not intended – score. As a matter of fact these misleading aspects were discovered after having carried out the IOR testing with two members of the DS team as one team member scored according to the guidelines (weight bearing=1, present). After having

[1]Weight shifting = If the donkey is standing look if it is shifting its weight from one leg to the other not wanting to bear weight on one or more legs too long (Note: this is usually of importance when the donkey is loaded.).
Weight bearing = The donkey is bearing its weight evenly on all four legs

inquired in which way he scored the lameness parameters his scores could be corrected before carrying out the statistical analysis. A corrected version of the guidelines is presented in the annex (p. 79).

Regarding the structure the final version of the protocol (see Annex, p. 79) starts with a **general part** which provides information about the observer, the date as well as the name of the site (stating also if it is a market, a village or an alternative location). Furthermore, it is noted if the animal is a donkey or a mule as well as how it is used (draft/pack/ridden/other). Then the sex is stated (stallion/gelding/mare).

The actual assessment starts with the assessment of **behaviour and demeanour** (representing the thumb of the Hands-on Donkey Welfare assessment protocol). The following aspects are assessed: general demeanour and reaction towards humans with regard to observer approach, chin and ear contact test and additionally tail tuck. The definitions were mainly taken from the protocol developed by *The Brooke* and *University of Bristol* (PRITCHARD and WHAY, not stated, PRITCHARD et al., 2005) and its further work based on that protocol (MCLEAN et al., 2012). Parameters were – if necessary – modified (e.g. chin and ear contact test) by changing or increasing the number of possible scoring units based on the knowledge and advice of the Ethiopian DS team. General demeanour was categorised into “Alert”, “Apathetic” and “Severely depressed”. The observer approach included five categories: “Moves away”, “Turns head away”, “Does not move”, “Turns head towards observer” and “Aggressive”. The categories of the chin contact test included: “Accepts chin contact”, “Avoids chin contact or withdraws head when chin contacted (i.e. touched)”, “Exaggerated avoidance of chin contact”. The ear contact test consisted of “Tolerates ear touch” (which also includes allowing the ear to be touched) and “Avoids ear touch”. According to the Ethiopian DS chief of staff donkeys generally avoid being touched at their ears but some severely depressed animals might not react to this contact test at all. Therefore the obtained results could give additional information in so far if the observed demeanour relates to the reaction shown during the ear contact. The third category “Exaggerated avoidance of chin contact” was chosen in order to get more detailed information on the donkeys’ reaction. After having tested the protocol, the classifications categories “total refusal of chin contact” and “total refusal of ear contact” and respective definitions were added “total refusal of chin contact” and “total refusal of ear contact” as it was not possible to assess a donkey’s reaction when it was moving away during the approach test.

The index finger (**body condition**) combines the following parameters: body condition scoring and teeth assessment (i.e. quidding, missing teeth, abscess, hooks). The assessment of body condition is based on the *The Donkey Sanctuary*’s body condition score card (THE DONKEY SANCTUARY, not stated-a) however without assigning half scores. The

categories are defined as follows: poor (1), moderate (2), ideal (3), fat (4) and obese (5). As the age is estimated by examining the teeth an approximate age assessment was done during the prior assessment of the teeth.

Lameness represented by the middle finger is covered by the following parameters: weight shifting, non weight bearing, limb abnormal, hoof abnormal, signs of inflammation on the two right limbs, sole surface of the two right hooves normal. It is recorded if the donkey is not “lame”, “potentially lame” or “obviously lame”. A donkey is scored as “potentially lame” when the donkey shows one of the problems stated above but does not or cannot move because it is tethered and cannot be examined during walking. For this reason, it is also recorded if the donkey is loaded or tethered. According to the experience of the Ethiopian DS team any problem assessed at the limb, the hoof etc. must be considered as a potential cause for lameness and pain. If lameness can be clearly assessed in a moving donkey then the donkey is scored as “obviously lame”. If lameness cannot be assessed during walking but the donkey shows one of the signs above then it is still scored as “potentially lame”,

The assessment of **wounds** (ring finger) is done by using body mapping together with a sketch of a donkey (left- and right side). Wound depth has four categories:

1 – hairless, old wounds, scars

2 – superficial, pale pink, abrasion

3 – skin and immediate subcutaneous layers broken

4 – deep lesion. In this case, a muscle, a tendon or a bone is visible.

Wound size was defined as A (up to two fingers in length i.e. index and middle finger), B (from two fingers to palm size) and C (bigger than palm size). The following procedure is applied to assess wounds in the context of body mapping: whenever a wound is assessed both wound depth and size are scored and the two characters (e.g. A for wound size, 2 for superficial wound) are assigned to the respective region on the image of the donkey (Figure 5).

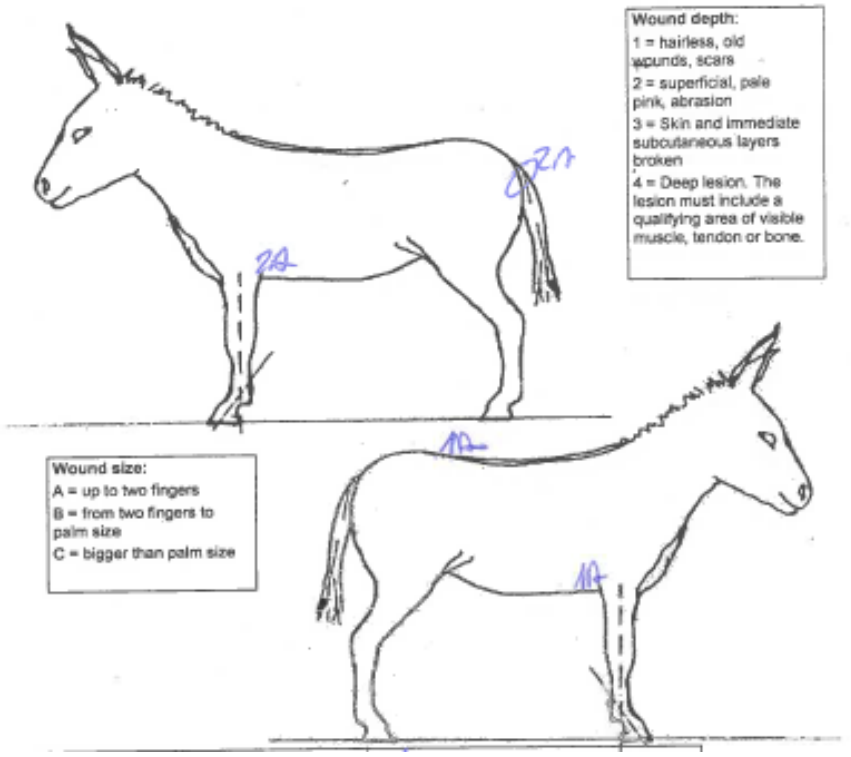


Figure 5: Example for application of body mapping during assessment of wounds

The little finger stands for **other signs of injury and disease** and includes the following parameters: coat condition, evidence of ectoparasites (plus stating the type of ectoparasite), heat stress, discharge from nose and eyes and other signs of diseases.

5.2 Donkey Welfare Assessment

General information on the donkeys

A total of 107 donkeys were assessed (SD: +/- 3). All of these animals were used as pack donkeys. The donkeys of all 11 different locations are described together assuming, that the use as pack animal is the same across locations.

The majority of assessed donkeys (n=101) were female (63.4 %), followed by 31.7 % stallions and 5 % geldings. Regarding the age distribution more than half of the assessed donkeys were older than 10 years Donkeys of an age of 15 and older represented 28 % of the assessed animals (Figure 6).

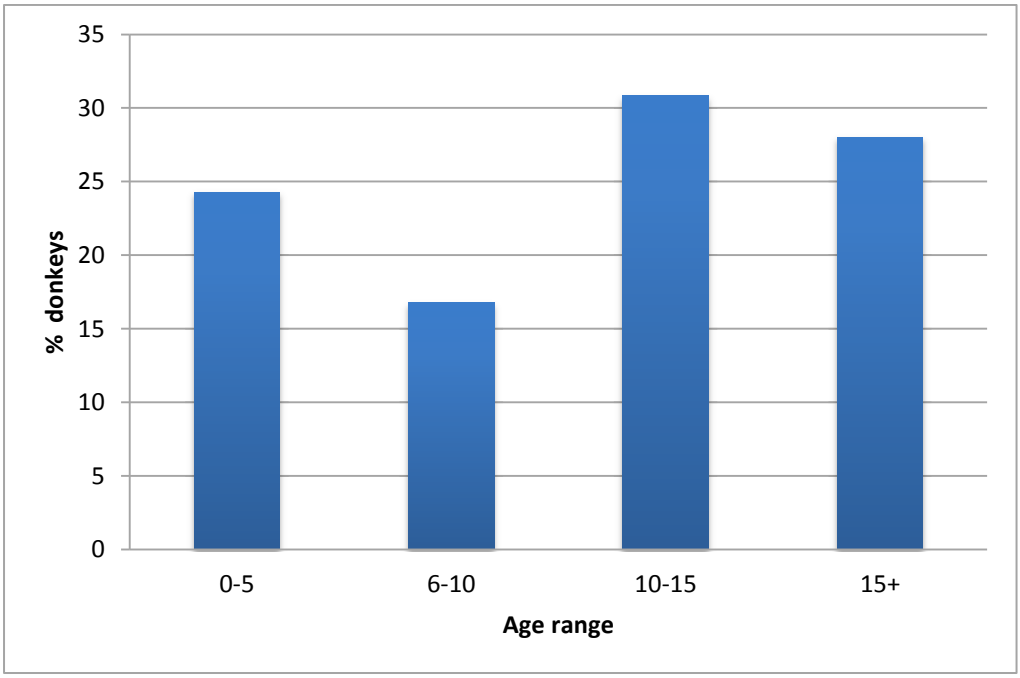


Figure 6: Age distribution of assessed donkeys (n=107)

Finger 1 (Thumb) / General demeanour

The behavioural assessment included the general demeanour, the approach of the observer and chin and ear contact test. Furthermore, it was tested if the donkey tucks the tail (Table 1). More than half of the assessed donkeys displayed alertness during the general demeanour assessment. 50 % showed an avoidance reaction towards the observer approach with either moving away or turning the head away. Only 4.7 % donkeys turned the head towards the observer; the rest (44.3 %) did not move. The majority of the donkeys avoided chin and ear contact. The percentage of donkeys that refused chin and ear contact totally (i.e. moving away during observer approach and chon contact test respectively) was high with 29.2 % and 47.7 %.

Table 1: Behavioural assessment of pack donkeys (n=107, SD: +/- 3)

	Donkeys (%)	Donkeys assessed (n)
General demeanour		107
Alert	54.2	
Apathetic	45.8	
Severely depressed	0	
Observer approach		106

	Donkeys (%)	Donkeys assessed (n)
Tail tuck	11	91
Chin contact		106
Accepts chin contact	31.1	
Avoids chin contact or withdraws head when chin contacted	34.9	
Exaggerated avoidance of chin contact	4.7	

	Donkeys (%)	Donkeys assessed (n)
Moves away	31.1	
Observer approach		
Does not move	44.3	
Turns head towards observer	4.7	
Aggressive	0.9	

	Donkeys (%)	Donkeys assessed (n)
Total refusal of chin contact	29.2	
Ear contact test		107
Tolerates ear contact	23.4	
Avoids ear contact	29.0	
Total refusal of ear contact	47.7	

Finger 2 (Index finger) / body condition

Nearly 82 % of the assessed animals had a **BCS** score of 2 (moderate). The proportion of donkeys with a poor BCS (1) was 16.5 % whereas only 2 donkeys (1.9 %) were scored as having an appropriate body condition (3).

Regarding teeth condition 74.8 % of 104 assessed donkeys had no apparent teeth problems. Detected teeth problems were deceased and missing teeth (5.7 %), calculus (4.8 %), abnormal position of jaw (4.8 %), to a lesser extent strange sound (when rubbing upper and lower jaw against each other), abnormal position of teeth and hooks.

Finger 3 (Middle finger)/ lameness

Regarding **lameness**, 38.3 % of the animals assessed were “not lame” whereas nearly 60 % were considered as “potentially lame” as problems regarding their limbs and hooves were present and therefore could cause subtle lameness which is difficult to assess in donkeys (see p. 30). 1.9 % of the donkeys were “obvious lame” as lameness could be clearly assessed during walking. Abnormal limbs which show for instance a not properly healed fracture were assessed in 28 % of 107 assessed donkeys whereas hoof abnormalities (e.g. cracked hoof, hoof too long) showed a prevalence of 31.8 % (Table 2)

Photo 7: Example for abnormal limb

Regarding signs of inflammation and abnormal surface on the right legs it cannot be tracked back if the abnormal surface was detected on the front or the back leg when both legs were assessed. No data are available of 7 and 17 donkeys respectively as the donkeys were for instance unwilling to have the hooves picked up. As specific results of either assessment of right front leg only or of both right legs do not give any additional information only the overall scores are presented in Table 2.

Table 2: Lameness assessment of pack donkeys

	Donkeys (%)	n
Weight shifting	0	107
Non weight bearing	4.7	107
Limb abnormal	28	107
Hoof abnormal	31.8	107
Signs of inflammation on one leg	3.2	87
Sole surface of one right hoof abnormal	36.4	83
Lame		107
Not lame	38.3	
Potentially lame	59.8	
Obviously lame	1.9	

Finger 4 (Ring finger) / wounds

In total, 94 (87.9 %) of the assessed donkeys had or one or more **wounds** have at the time of the assessment (Photo 8). More than 68 % had suffered or suffered from more than one wound (Ø 2.9 wounds per donkey).

Photo 8: Donkey with wounds scored 1A

Of those 94 donkeys nearly 75 % had one or more wounds classified as 1A (hairless, old wounds, scars up to the size of two fingers). 90.5 % of donkeys with wounds had one or more wounds with the size of A (up to the size of two fingers).

Regarding wound depth Figure 7 shows the prevalence of different wound depths on donkeys with wounds (n= 94) whereas Figure 8 displays the prevalence of different wounds regarding size.

Of those 94 donkeys having a wound 73.4 % had a back wound. 36.2 % and 20.2 % had wounds on the front and/or hind legs respectively. The so called predisposed area around the breast (i.e. defined by the author and the DS team as area which is prone to wounds) was a problem with 31.9 % of the donkeys (rest of breast and belly: 4.3 %); 39.4 % had a wound under the tail. Head wounds were found in 18.1 % of the donkeys.

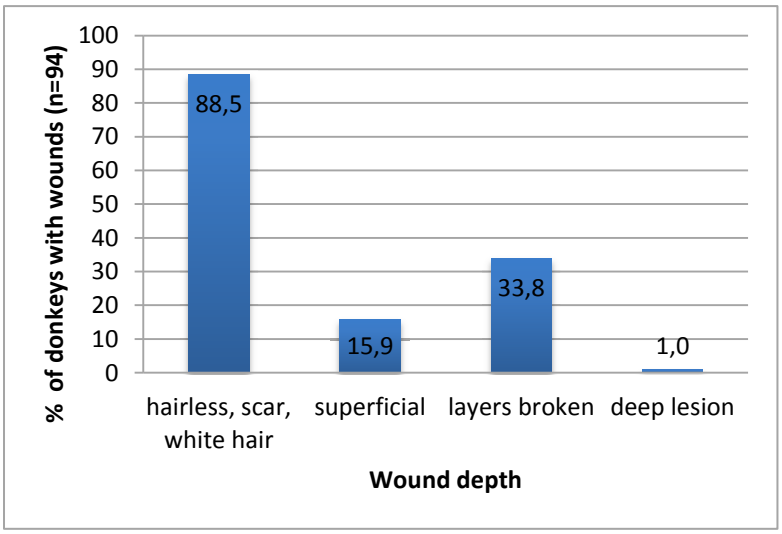


Figure 7: Prevalence of wounds regarding wound depth

The evidence of a hyena bite was seen in 6 donkeys or 6.1 % respectively. Ear lesions were summarized into three categories: no problem (88.8 %), 5.6 % with a wound or wounds with the score 1A/2A/2B and 1.9 % with a wound or wounds with 3/4 A/B/C. 3.7 % had an ear partly bitten off or “missing”.

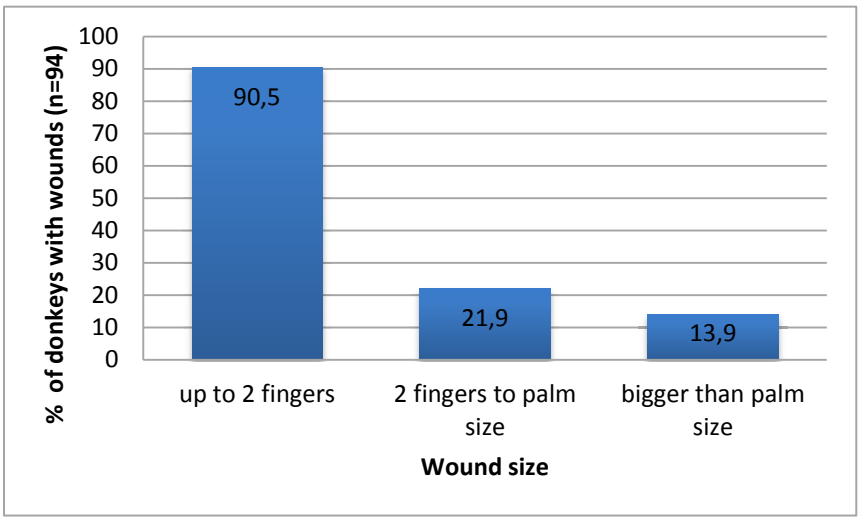


Figure 8: Prevalence of wounds regarding wound size

Finger 5 (Little finger) / Other signs of injury and disease

Regarding **other signs of injury and disease**, dirtiness and wetness of the coat and the legs was noted: 77.6 % of the donkeys were clean, with 10.3 % had a wet coat and 11.2 % a dirty coat or legs. One donkey was both dirty and wet. This aspect was noted in order to point out potential difficulties when assessing the coat condition as well as signs of inflammation.

Ectoparasite infestation was seen in 41.1 % of the assessed donkeys (n=107) with 2.8 % of the animals being heavily infested (Table 3, also displaying the results of the assessment of coat condition).

Table 3: Coat conditions and infestation with ectoparasites (n=107)

	Donkeys (%)
Coat condition	
Excellent	6.6
Average	67.9
Poor	25.5
Infestation with ectoparasites	
No infestation	58.9
Little to moderate infestation	38.3
Heavy infestation	2.8

The most prevalent ectoparasites found on the donkeys were Gasterophilus (eggs on mane and breast) with 31.1 % followed by signs of mud fever (Photo 9) and flies infestation (9.3 %) and symptoms indicating mange (6.5 %). Ticks were seen on one donkey.

Photo 9: Signs of mud fever

One out of 106 donkeys showed signs of heat stress.

Nose and eyes discharge was found in 43 % and 41 % of the animals respectively (Figure 9).

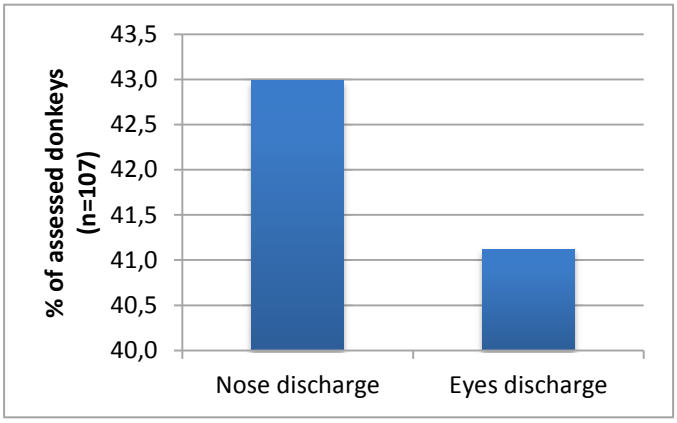


Figure 9: Nose and eyes discharge (n=107)

Other signs of diseases were written down separately and the following welfare problems occurred: 9.3 % of the donkeys had a swelling (or swellings) on the back whereas other aspects such as eye problems (blindness), part of tail missing or deformed or slit nostrils appeared to a much lesser extent (i.e. on one donkey).

5.3 Inter-Observer Reliability

Since none of the observers was considered a Gold Standard, the three observers were compared to each other.

Whenever one observer rated as a constant (i.e. an observer always scored the same category e.g. heat stress absent in all observed donkeys) it was not possible to calculate Kappa in SPSS. Mathematically speaking the result was 0 – however this does not mean that the k-value was 0 (i.e. poor agreement). If this occurred in the present dataset no results are provided with stating that one or more observer(s) are constants (Table 4). In several variables constant ratings occur such as “Sex: gelding” or “Chin contact: total refusal”.

For specific ordinal parameters such as “Lame” or “Coat condition” the tool used (at http://vassarstats.net/kappa.html) did not compute k-values as the observed concordance was smaller than the mean-chance concordance.

Furthermore, the Kappa value as well as the weighted Kappa is presented along with the percentage agreement and the PI and BI. The PABAK gives additional information as it shows the effect of the PI and BI on the Kappa values.

The PA ranges between 41 % to 100 % in the pair-wise comparison of Observer 1 and Observer 2. The PI and the BI lie between 0.13 and 0.91 and 0 and 0.41 respectively. The Kappa ranges from -0.04 to 1 whereas the PABAK lies between -0.06 and 1.

In the second pair-wise comparison the values range in the following way: PA 44 % to 100%, PI 0.03 to 0.91, BI 0 to 0.41, Kappa -0.06 to 0.91 and the PABAK -0.13 to 1.

Regarding the agreement of Observer 2 and Observer 3 the PA lies between 34 % and 100 %. The PI and the BI range from 0.28 to 0.94 and from 0 to 0.34 respectively, with a Kappa between -0.28 and 1. The PABAK lies between -0.31 and 1.

Following BURN et al. (2009) a PI of less than 0.25 indicates a well-balanced population. According to this threshold, only three parameters ("Observer approach: Does not move", "Chin contact: accept", "Ear test: tolerate") can be regarded as balanced when comparing Observer 1 and 2. In the other two pair-wise comparisons three and two parameters were found to have a balanced population, respectively.

On the other hand the prevalence index was high (≥ 0.75) for five, eight and ten parameters in the corresponding pair-wise comparisons. This indicates a highly unbalanced population. Those parameters whose k-value could not be obtained because of constant rating – 12, seven and ten in the respective pair-wise comparison – also indicate an unbalanced population: the observers constantly rate the same category due to a constant (non-) occurrence of one category.

Table 4: Inter-Observer Reliability testing of a working donkey welfare assessment (n=32)

Observer 1 – Observer 2					
Variable	Percentage Agreement (%)	Prevalence Index (PI)	Bias Index (BI)	Kappa/Kendall's	PABAK
Sex	100%				1
Stallion	100%	0.33	0	1	1
Gelding	100%			Both observers constant	1
Mare	100%	0.33	0	1	1
General Demeanour[a]	63%			0.06	0.25
Observer Approach (OA)	47%				-0.06
OA Moves away	81%	0.56	0.06	0.46	0.63
OA Turns head away	59%	0.34	0.03	0.08	0.19
OA Does not move	56%	0.13	0.13	0.13	0.13
OA Turns head towards observer	100%			Both observers constant	1.00
OA Aggressive	100%			Both observers constant	1.00
Tail Tuck	77%	0.63	0.10	0.23	0.53
Chin Contact (CC)	63%				0.25
CC Accept	72%	0.22	0.16	0.42	0.44
CC Avoid	72%	0.47	0.03	0.28	0.44
CC Exaggerated Avoidance	97%	0.91	0.03	0.65	0.94
CC Total Refusal	84%			Observer 2 constantt	0.69
Ear Test (ET)	41%				-0.19
ET Tolerate	53%	0.16	0.16	0.06	0.06
ET Avoid	56%	0.44	0.00	-0.08	0.13
ET Total Refusal	72%	0.72	0.16	-0.11	0.44
BCS[a]	72%			0.29	0.44
Teeth	83%			Observer 2 constant	0.67
Age[a]	47 %			0.29	-0.06
Weight Shifting	91%			Observer 1 constant	0.81
Non Weight Bearing	88%	0.81	0.06	0.27	0.75
Limb Abnormal	56%	0.50	0.44	0.07	0.13
Hoof Abnormal	66%	0.28	0.09	0.26	0.31
Signs Inflammation Front Leg	100%			Both observers constants	1

Observer 1 – Observer 2					
Variable	Percentage Agreement (%)	Prevalence Index (PI)	Bias Index (BI)	Kappa/Kendall's	PABAK
Sole Surface Front Leg	77%			Observer 2 constant	0.55
Lame[a]	56%			Observed concordance smaller than mean-chance concordance	0.13
Wounds Head Count	97%	0.91	0.03	0.65	0.94
Wounds Neck Count	97%			Observer 2 constant	0.94
Wounds Back Count	88%	0.44	0.13	0.70	0.75
Wounds Hip Count	94%			Observer 2 constant	0.88
Wound Under Tail Count	68%	0.39	0.32	0.20	0.35
Wound Hind Leg Count	75%	0.50	0.06	0.34	0.50
Wound Front Leg Count	63%	0.50	0.25	0.08	0.25
Wound Breast Predisposed Count	69%	0.63	0.31	0.12	0.38
Wound Rest Breast Belly Count	91%	0.91	0.03	-0.04	0.81
Coat Condition[a]	53 %			Observed concordance smaller than mean-chance concordance	0.06
Evidence Ectoparasites[a]	72 %			0.55	0.44
Heat stress	97 %			Observer 2 constant	0.94
Discharge Nose	52 %	0.32	0.35	0.05	0.03
Discharge Eyes	58 %	0.52	0.10	0.08	0.16

[a] Weighted Kappa value

Observer 1 – Observer 3					
Variable	Percentage Agreement (%)	Prevalence Index (PI)	Bias Index (BI)	Kappa/Kendall's	PABAK
Sex	94%				0.87
Stallion	94%	0.35	0.06	0.85	0.88
Gelding	94%				0.88
Mare	97%	0.32	0.03	0.93	0.94
General Demeanour[a]	56%			0.17	0.13
Observer Approach (OA)	45%				-0.10
OA Moves away	84%	0.59	0.03	0.52	0.68
OA Turns head away	48%	0.09	0.38	0.01	-0.03
OA Does not move	58%	0.34	0.34	0.19	0.16
OA Turns head towards observer	100%			Both observers constants	1.00
OA Aggressive	100%			Both observers constants	1.00
Tail Tuck	77%	0.67	0.07	0.29	0.53
Chin Contact (CC)	47%				-0.06
CC Accept	53%	0.34	0.28	0.02	0.06
CC Avoid	66%	0.66	0.22	-0.11	0.31
CC Exaggerated Avoidance	94%	0.94	0.00	-0.03	0.88
CC Total Refusal[a]	81%	0.75	0.06	0.15	0.63
Ear Test (ET)	53%				0.06
ET Tolerate	66%	0.28	0.28	0.31	0.31
ET Avoid	69%	0.63	0.19	0.03	0.38
ET Total Refusal	72%	0.66	0.09	0.03	0.44
BCS[a]	75%			0.63	0.50
Teeth	100%			Both observers constants	1.00
Age[a]	58%			0.59	0.16
Weight Shifting	91%			Observer 1 constant	0.81
Non Weight Bearing	91%	0.84	0.03	0.35	0.81
Limb Abnormal	59%	0.41	0.34	0.15	0.19
Hoof Abnormal	57%	0.11	0.31	0.24	0.14
Signs Inflammation Front Leg	97%			Both observers constant	0.93
Sole Surface Front Leg	77%			Observer 3 constant	0.55
Lame[a]	44%			0.19	-0.13
Wounds Head Count	100%	0.94	0.00	1.00	1.00

Observer 1 – Observer 3					
Variable	Percentage Agreement (%)	Prevalence Index (PI)	Bias Index (BI)	Kappa/Kendall's	PABAK
Wounds Neck Count	97%			Observer 3 constant	0.94
Wounds Back Count	84%	0.41	0.16	0.64	0.69
Wounds Hip Count	94%	0.88	0.00	0.47	0.88
Wounds Under Tail Count	74%	0.48	0.26	0.38	0.48
Wounds Hind Leg Count	75%	0.75	0.19	-0.06	0.50
Wounds Front Leg Count	69%	0.31	0.13	0.29	0.38
Wounds Breast Predisposed Count	81%	0.50	0.19	0.52	0.63
Wounds Rest Breast Belly Count	91%	0.91	0.03	-0.04	0.81
Coat Condition[a]	50%			0.45	0.00
Evidence Ectoparasites[a]	72%			0.48	0.44
Heat stress	100%	0.94	0.00	1.00	1.00
Discharge Nose	66%	0.03	0.09	0.32	0.31
Discharge Eyes	53%	0.53	0.41	-0.06	0.06

[a] Weighted Kappa value

Observer 2 – Observer 3					
Variable	Percentage Agreement (%)	Prevalence Index (PI)	Bias Index (BI)	Kappa	PABAK
Sex	90%				0.81
Stallion	88%	0.39	0.10	0.78	0.75
Gelding	94%				0.88
Mare	97%	0.32	0.03	0.93	0.94
General Demeanour[a]	66%			0.39	0.31
Observer Approach (OA)	32%				-0.35
OA Moves away	84%	0.53	0.03	0.57	0.68
OA Turns head away	35%	0.06	0.34	-0.18	-0.29
OA Does not move	48%	0.47	0.22	-0.28	-0.03
OA Turns head towards observer	100%			Both observers constant	1.00
OA Aggressive	100%			Both observers constant	1.00
Tail Tuck	75%	0.63	0.06	0.27	0.50
Chin Contact (CC)	59%				0.19
CC Accept	69%	0.50	0.13	0.18	0.38
CC Avoid	69%	0.69	0.19	-0.11	0.38
CC Exaggerated Avoidance	91%	0.91	0.03	-0.04	0.81
CC Total Refusal	91%			Observer 2 constant	0.81
Ear Test (ET)	56%				0.13
ET Tolerate	63%	0.44	0.13	0.09	0.25
ET Avoid	63%	0.56	0.19	-0.16	0.25
ET Total Refusal	88%	0.81	0.06	0.27	0.75
BCS[a]	63%			0.47	0.25
Teeth	100%			Both observers constants	1.00
Age[a]	68%			0.65	0.35
Weight Shifting[a]	90%	0.84	0.03	0.35	0.81
Non Weight Bearing	84%	0.75	0.03	0.20	0.69
Limb Abnormal	84%	0.84	0.09	-0.05	0.69
Hoof Abnormal	53%	0.14	0.25	0.19	0.06
Signs Inflammation Front Leg	100%			Both observers constants	1
Sole Surface Front Leg	100%			Both observers constants	1
Lame[a]	56%			0.32	0.13
Wounds Head Count	97%	0.91	0.03	0.65	0.94
Wounds Neck Count	100%			Both observers constant	1.00

Observer 2 – Observer 3					
Variable	Percentage Agreement (%)	Prevalence Index (PI)	Bias Index (BI)	Kappa	PABAK
Wounds Back Count	84%	0.28	0.03	0.66	0.69
Wounds Hip Count[a]	94%			Observer 2 constant	0.88
Wounds Under Tail Count	81%	0.81	0.06	-0.09	0.63
Wounds Hind Leg Count	69%	0.69	0.25	-0.06	0.38
Wounds Front Leg Count	69%	0.63	0.13	0.00	0.38
Wounds Breast Predisposed Count	88%	0.81	0.13	0.30	0.75
Wounds Rest Breast Belly Count	100%	0.94	0.00	1.00	1.00
Coat Condition[a]	34%			Observed concordance smaller than mean-chance concordance	-0.31
Evidence Ectoparasites[a]	88%			0.67	0.75
Heat stress	97%			Observer 2 constant	0.94
Discharge Nose	55%	0.42	0.26	-0.01	0.10
Discharge Eyes	94%	0.94	0.00	-0.03	0.87

[a] Weighted Kappa value

Table 5 contains the assignment of the different parameters to the categories of reliability as suggested by BURN et al. (2009).

Disregarding the parameters whose k-values could not be obtained because of constant rating Table 5 shows that the vast majority of the remaining parameters showed poor agreement for all observer pairs. The parameter “Sex” showed substantial and excellent agreement respectively. Most of those parameters used to assess the behaviour such as “General demeanour” and “Observer approach” cannot be regarded as useful in terms of inter-observer agreement.

Some pairs had better agreement within specific parameters compared to the other pairings. For instance, poor agreement was received within the parameter “Age” for Observer 1 and 2, whereas there was moderate and substantial agreement within the remaining two pair-wise comparisons. The parameter “Evidence of ectoparasites” which included ordinal variables showed moderate (2 times) and substantial agreement whereas another example for a parameter consisting of ordinal variables – “Coat condition” – showed moderate agreement

within one pair-wise comparison. For the remaining two, the k-value was not obtained due to a smaller observed consistency than by mean-chance agreement.

Table 5: Reliability rating of a donkey welfare assessment (k-values and weighted k-values)

Observer 1 – Observer 2				
POOR	**AMBIGUOUS**	**MODERATE**	**SUBSTANTIAL**	**EXCELLENT**
PA < 75 % for binary variables **PA < 75 % and weighted k < 0.4 for ordinal variables**	**PA ≥ 75 % but k < 0.40** **PA ≥ 75 % but weighted k > 0.4 for ordinal variables**	**k = 0.40 – 0.59**	**k = 0.60 – 0.79**	**k = 0.80 – 1.00**
General Demeanour	Tail tuck	Observer Approach (Moves Away)	Chin contact (Exaggerated Avoidance)	Sex (Stallion)
Observer Approach (Turns head away)	Non Weight Bearing	Chin contact (Accept)	Wounds Head Count	Sex (Mare)
Observer Approach (Does not move)	Wound Hind Leg Count	Evidence Ectoparasites	Wounds Back Count	
Chin contact (Avoid)	Wound Breast Predisposed Count			
Ear test (Tolerate)				
Ear test (Avoid)				
Ear test (Total refusal)				
Age				
Limb Abnormal				
Hoof Abnormal				
Wound Under Tail Count				
Wound Front Leg Count				
BCS				
Heat stress				
Discharge Nose				
Discharge Eyes				

Observer 1 – Observer 3				
POOR	**AMBIGUOUS**	**MODERATE**	**SUBSTANTIAL**	**EXCELLENT**
PA < 75 % for binary variables **PA < 75 % and weighted k < 0.4 for ordinal variables**	**PA ≥ 75 % but k < 0.40** **PA ≥ 75 % but weighted k > 0.4 for ordinal variables**	**k = 0.40 – 0.59**	**k = 0.60 – 0.79**	**k = 0.80 – 1.00**
General Demeanour	Tail tuck	Age	Wounds Back Count	Sex (Stallion)
Observer Approach (Turns head away)	Chin contact (Exaggerated Avoidance)	Observer Approach (Moves away)	BCS	Sex (Mare)
Observer approach (Does not move)	Chin contact (Total refusal)	Wounds Hip Count		Wounds Head Count
Chin contact (Accept)	Non Weight Bearing	Wounds Breat Predisposed Count		Heat stress
Chin contact (Avoid)	Wounds Front Leg Count	Evidence Ectoparasites		
Ear Test (Tolerate)	Wounds Rest Breast Belly Count	Coat Condition		
Ear Test (Avoid)				
Ear Test (Total refusal)				
Limb Abnormal				
Hoof Abnormal				
Lame				
Wounds Under Tail Count				
Discharge Nose				
Discharge Eyes				

Observer 2 – Observer 3				
POOR	**AMBIGUOUS**	**MODERATE**	**SUBSTANTIAL**	**EXCELLENT**
PA < 75 % for binary variables **PA < 75 % and weighted k < 0.4 for ordinal variables**	**PA ≥ 75 % but k < 0.40** **PA ≥ 75 % but weighted k > 0.4 for ordinal variables**	**k = 0.40 – 0.59**	**k = 0.60 – 0.79**	**k = 0.80 – 1.00**
General Demeanour	Tail tuck	Observer Approach (Moves away)	Sex (Stallion)	Sex (Mare)
Observer Approach (Turns head away)	Chin contact (Exaggerated avoidance)	BCS	Age	Wounds Rest Belly Count
Observer Approach (Does not move)	Weight Shifting		Wounds Head Count	
Chin contact (Accept)	Non Weight Bearing		Wounds Back Count	
Chin contact (Avoid)	Limb abnormal		Evidence Ectoparasites	
Ear Test (Tolerate)	Wounds Under Tail Count			
Ear Test (Avoid)	Wounds Breast Predisposed Count			
Ear Test (Total Refusal)	Discharge Eyes			
Hoof abnormal				
Lame				
Wounds Hind Leg Count				
Wounds Front Leg Count				
Discharge Nose				

5.4 Qualitative Behaviour Assessment (QBA)

Several descriptive terms were generated during the four observations of donkey groups. In the beginning, the terms were primarily described what the donkeys do (e.g. flapping with the ears. tail swishing). However in the course of the process the observers collated several terms which describe how the donkeys behave. After Phase 1, the following terms were collated: **depressed, active, friendly, exhausted, curious, feeling discomfort, social, relaxed, happy**. In Phase 2, in the first line these terms were applied to individual donkeys.

The results of the overall analysis are the following: Principal Component Analysis extracted three components with eigenvalues greater than one which were further considered. The first component explains 44.1 %, the second 27.2 % and the third 11.5 % of the variance. Table 6 shows which term loads more than +/- 0.6 on each principal component.

The **first axis** includes terms ranging from **negative** (depressed, exhausted) **to positive descriptors** of mood or mental state (active, happy, curious). These components could be summarized as “General mental state”. The second axis consists of “social” and “friendly” and has no negative dimension. It could be described as “Responsiveness towards social stimuli”. The third main component consists solely of the term “Relaxed”.

Table 6: Terms and loadings describing principal components 1, 2 and 3 (terms loading more than +/-0.6 are recorded)

	Component		
	1	**2**	**3**
	General feeling	**Reaction towards environment**	**Relaxed**
Depressed	-0,873		
Active	0,832		
Friendly		0,797	
Exhausted	-0,71		
Curious	0,737		
FeelingDiscomfort		0,661	
Social		0,901	
Relaxed			0,824
Happy	0,934		

Figure 10 shows the plot of Component 1 **“General mental state”** against Component 2 **“Responsiveness towards social stimuli”**. Firstly, it displays the negative and positive terms for the first axis. Secondly, the term “Feeling discomfort” is considered as a descriptive

part of Component 2. However, there is also a proximity to the terms “exhausted” and “depressed” (i.e. the negative end of Component 1).

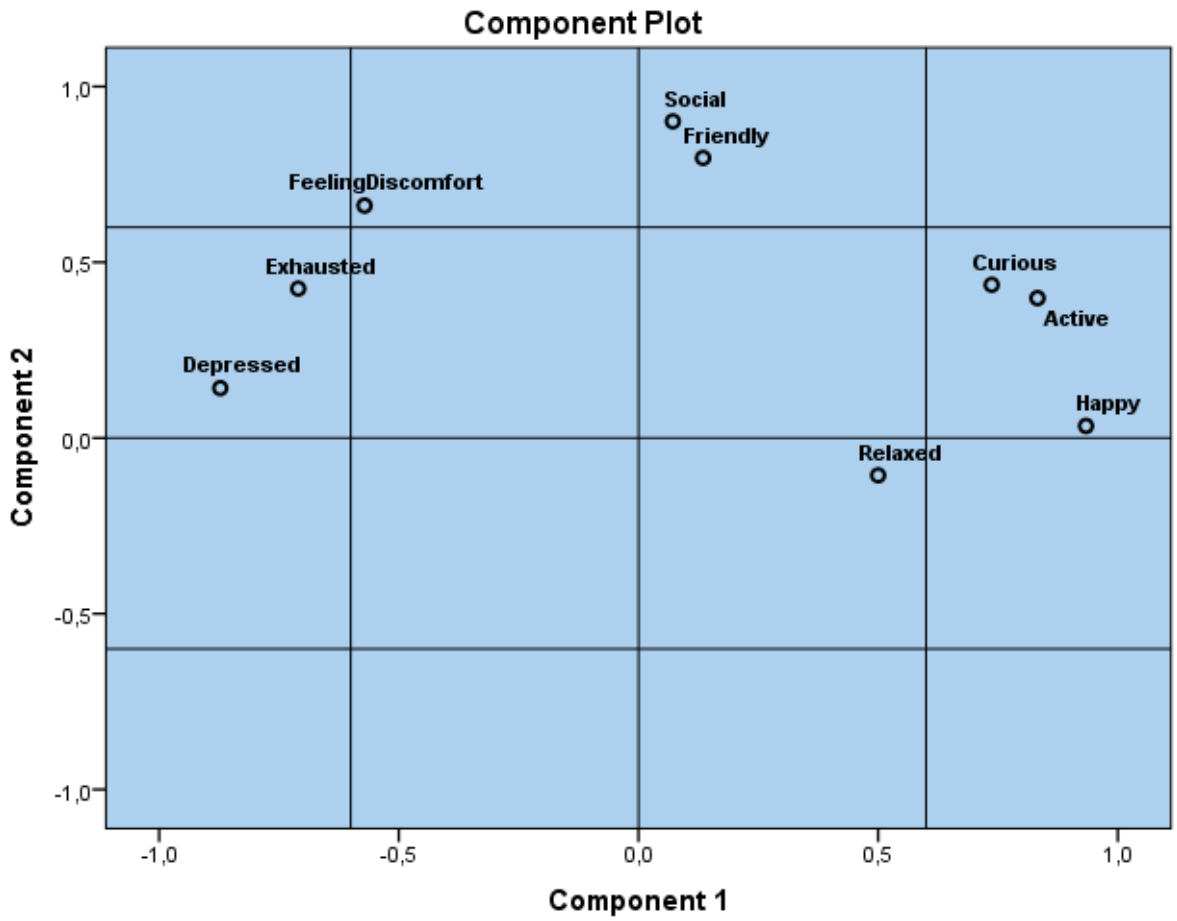


Figure 10: Plot with main Components 1 and 2

For a better overview of the complexity regarding each main component Table 7 shows loadings resulting from PCAs which were carried out separately for each of the three observers as compared to the overall analysis. Independent of the analysis, component 1 was defined by the terms **active**, **curious**, **happy**, **depressed** and **exhausted**. Regarding the remaining terms there is an inconsistency regarding the loadings gained from the different individual analyses. For instance, “Feeling discomfort” is displayed in the second component in the overall analysis and in the analysis of Observer 1 whereas the loadings for this term are highest in the first component in the case of Observer 2 and 3. The overall analysis displayed “Relaxed” as a main component which however has high loadings in the first component for Observer 2 and 3. This indicates that the observers had a general common understanding of some terms but that there is a difference in the individuals’ understanding of specific terms.

Table 7: Terms and loadings comparing overall analysis and individual observers

Components	Overall analysis	Obs 1	Obs 2	Obs 3
First component				
	Active (0.832)	Active (0.923)	Active (0.904)	Active (0.924)
	Curious (0.737)	Curious (0.934)	Curious (0.658)	Curious (0.929)
	Happy (0.934)	Happy (0.906)	Happy (0.906)	Happy (0.952)
		Friendly (0.714)	Relaxed (0.909)	Relaxed (0.949)
	Depressed (-0.873)	Depressed (-0.937)	Depressed (-0.874)	Depressed (-0.841)
	Exhausted (-0.710)	Exhausted (-0.880)	Exhausted (-0.744)	Exhausted (-0.855)
			Feeling discomfort (-0.865)	Feeling discomfort (-0.742)
Second component				
	Feeling Discomfort (0.661)	Feeling discomfort (0.809)		
	Friendly (0.797)		Friendly (0.967)	Friendly (0.920)
	Social (0.901)			Social (0.902)
		Relaxed (-0.568)		
Third component				
	Relaxed (0.824)			
		Social (0.933)	Social (0.860)	

Figure 11 displays the scoring of each individual donkey per observer based on the overall PCA. It shows that the three observers scored differently regarding the second dimension (see Observer 1 and 3 compared to Observer 2).

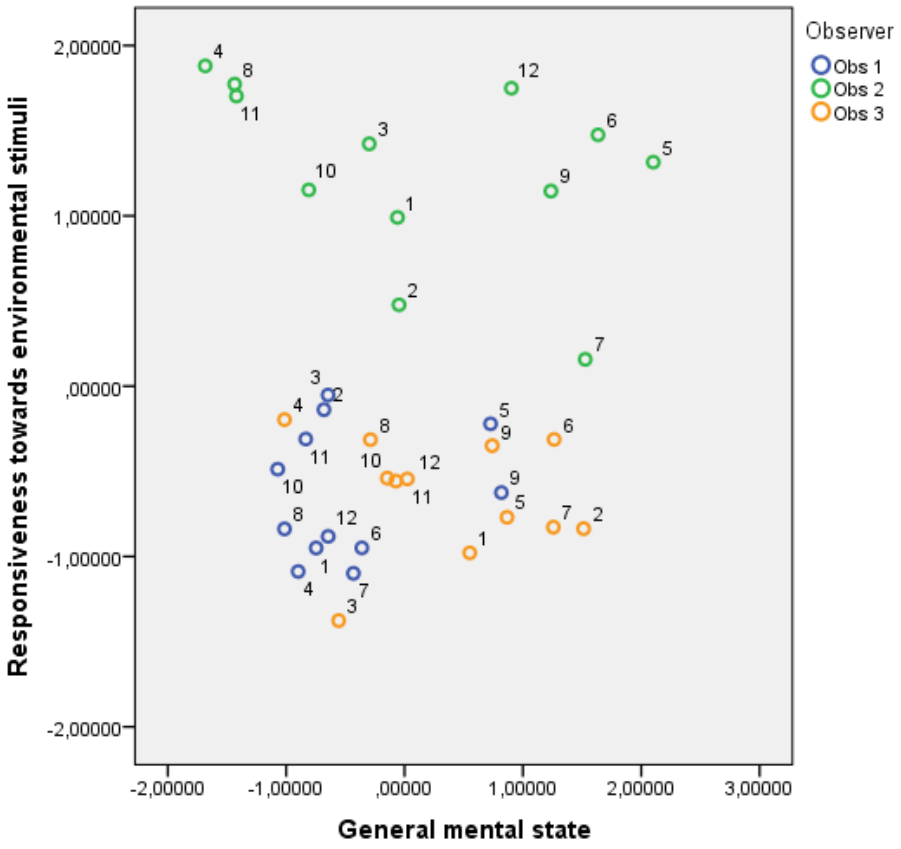


Figure 11: Regression factor scores for each donkey per observer

The **individual assessments** of the first component "General mental state" are displayed in Figure 12, focusing on individual donkeys.

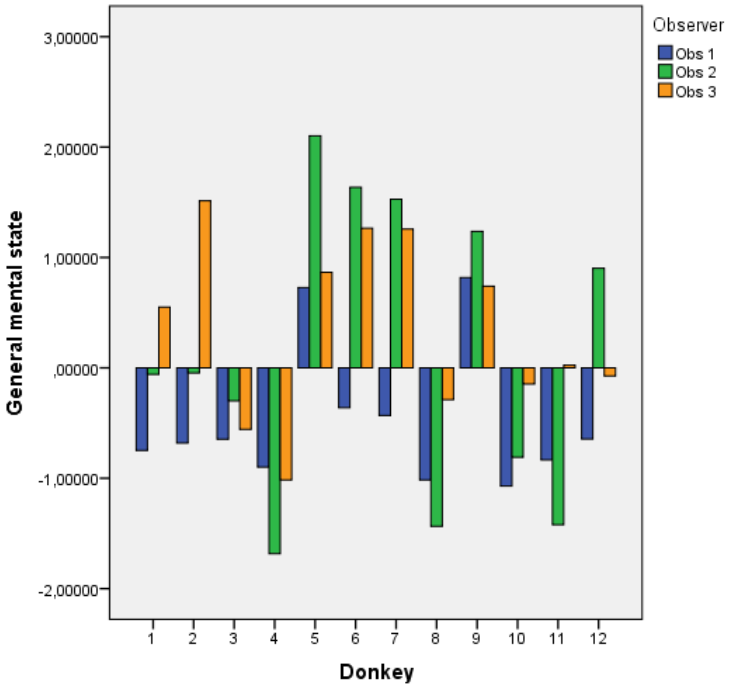


Figure 12: Individual loading of first main component (per donkey)

It shows that the observers judged six donkeys similarly (above or below neutral) regarding the first component “General mental state” (donkey 3. 4. 5. 8. 9. 10) whereas there is no common pattern in the assessment of the remaining six animals.

In a last round three team members of DS and the author observed a group of grazing donkeys near Debre Zeit. The data derived from this observation not statistically analysed but it served to discuss the QBA approach. During the discussion among the DS team members and the author, the terms “Friendly”. “Curious” and “Feeling discomfort” were the main topics of the discussion. Looking at the raw data (i.e. where the observers ticked the scale) these three terms showed a huge inconsistency between the three DS team members and the author which – at least with the term “Feeling discomfort” – corresponds to the results presented above. The discussion continued to define what each of the terms includes. The term “Friendly” involved the friendliness of the donkeys towards another donkey and towards a human. Checking when another animals or a human is approaching, the observation of movement and the interest in distant movements are part of the term “Curious”. A donkey that exhibits “Feeling discomfort” may be tethered at one place; additionally flies may contribute to “Feeling discomfort”. The discussion included to talk about the donkey expressing its feelings and imagining how, for instance, the happiest and unhappiest donkey would look like. Regarding the last point the different background between the three Ethiopian team members and the author regarding donkeys became more obvious as the Ethiopians would score these grazing donkeys as maximal happy along the VSA whereas the author scored these donkeys lower. While the Ethiopian team members know donkeys as working animals the author has her background mainly with donkeys kept as companion animals. In summary this last discussion round revealed that differences in the perception of working donkeys’ body language ranged from problems with the application of QB to various individual backgrounds (cultural, professional).

6 Discussion

In the following two paragraphs the developed Hands-on Donkey Welfare assessment protocol and the results obtained from its application on the field as well as the QBA-results are discussed. Though separately discussed, there are overlaps between these two chapters due to the to some extent continuous process of development, application and modification of the protocol.

6.1 The Extended Hands-on Donkey Welfare Assessment Protocol

The final protocol, which was applied in the field in Ethiopia, was a combination of already existing parameters for equine welfare assessment (e.g. general demeanour) as well as newly developed parameters (e.g. lameness parameters. body mapping for wounds assessment). Corresponding to the original version of the Hands-on Donkey Welfare assessment protocol the five fingers **behaviour/demeanour**, **BCS**, **lameness**, **wounds** and **other signs of injury and disease** were represented by specific parameters and assessment tools. Parameters were either chosen from other welfare assessment protocols such as the guidance notes for working equine welfare assessment (PRITCHARD and WHAY, not stated, PRITCHARD et al., 2005) and their further developments and alterations (e.g. MCLEAN et al., 2012) or based on the DS experiences (e.g. body mapping). The protocol had been pilot tested several times and changes had to be made before it was finalised. These alterations are also discussed below.

The protocol starts with a general part followed by parameters assessing the **behaviour and demeanour** (thumb). The parameters observer approach, the chin contact test and the ear contact test were slightly altered in order to make concise distinctions regarding the behaviour of the donkeys towards humans. The DS team and the author decided to name one newly defined category “Exaggerated avoidance” in the chin contact test. However, “exaggerated” might actually be a judgemental assessment of the donkey’s behaviour as the animal itself could see its reaction appropriate to the approach of a human. Therefore, it might be more suitable to change “Exaggerated” into “Strong avoidance”. During the pilot testing the category “Total refusal” in the chin and ear contact test was added to note if the donkey was refusing to be touched at all as it was moving away. It was considered important to note the difference between avoidance behaviour during these tests and a complete refusal. Logically, a donkey that is moving away during the observer approach shows avoidance during the contact tests too. However, in order to make distinctions between those donkeys that do not move away during the approach test and show avoidance behaviour

afterwards and those that refused in general this distinction is considered to be valuable and important.

The **body condition assessment** (index finger) consists of parameters which are considered to detect the body condition as well as aspects of nutrition and (impaired) energy input: BCS and teeth condition. The assessment of the body condition was carried out according to the chart and guidelines of DS. These two tools have been used regularly in the working process of the DS. However, the half-scores were omitted following the recommendations of the Director of International Relations Stephen Blakeway. The assessment of the body condition must be considered fundamental as a poor body condition is not only a welfare problem itself but also indicates towards several other welfare issues such as gait and sole abnormalities (BURN et al., 2010b). The body condition scoring is followed by the teeth assessment Teeth condition is considered to correlate to the body condition (DU TOIT et al., 2008, CYRIL ROY, 2006). As the age was estimated by checking the teeth this parameter was combined with BCS and teeth. Teeth problems detected during the assessment were more variable than the four categories provided in the protocol. Two of the problems detected but not covered by the protocol were diastemata and calculus. DU TOIT and DIXON (2012) report that diastemata (space between teeth in the jaw) can cause oral pain – however, closed diastemata are considered to be more welfare relevant as food can be trapped in the gap. This type of diastemata was not detected in the Ethiopian donkeys but only open diastemata on the front teeth which less likely lead to periodontal diseases. In extreme cases, calculus (mostly seen in canine teeth) can lead to gingivitis and periodontal diseases. This must be taken into account in a future donkey welfare assessment. However, the teeth assessment carried out in Ethiopia was insofar limited as only the front teeth could be visually assessed. The molar teeth were checked by running the hands up and down the molars and then by grinding the upper and lower jaw. DU TOIT et al. (2008) carried out detailed clinical dental examinations with 203 working donkeys in Mexico and found 18 % of these animals to have dental diseases which were highly welfare relevant and required instant medical care. These diseases included diastemata, overgrown teeth, worn teeth etc. All in all, 62 % of the observed donkeys suffered from diseases which included small focal overgrowths and soft tissue injuries. These reported numbers are higher than the detected teeth problems in this study (prevalence 25.2 %) and show that teeth problems are indeed a serious welfare issue. However, during on-field welfare assessment it is not feasible to carry out a detailed clinical examination. As the present results are not as precise compared to DU TOIT et al. (2008) a clear recommendation which teeth problems to put in focus cannot be given. Bearing the studies of DU TOIT et al. (2008) and DU TOIT and DIXON (2012) in mind that overgrown/worn and missing teeth likely lead to pain-related problems it is advisable to examine to what extent these serious welfare issues can be detected by using the above

described technique. Bearing the scientific literature in mind the problems detected – open diastemata and calculus – less likely lead to impaired health and welfare. If extreme cases occur they are accompanied by gingivitis and other periodontal problems which then must be regarded as a welfare problem. Therefore open diastemata and calculus should not be included in a list of detected teeth problems.

For the **lameness assessment** (middle finger) several conditions indicating problems were discussed. During the assessment the abnormal conditions should be clearly visible. It proved to be quite tricky in the field to assess these abnormal conditions properly as the assessment took place in markets and mill houses. Donkeys used to be tethered wherever possible and therefore they might not stand in their normal position. Thus the observer had to make sure to move them and look at the conditions several times from different angles. A good example for this necessary careful assessment is the condition “cow hocked” especially because it was mentioned by one DS team member that donkeys also show a slightly cow-hocked position when they are tired. This must be further taken into account in the next development stages of the protocol.

Furthermore, the parameters focusing on lameness in the first drafts of the protocol were going into more detail. This was changed in the final version of the protocol which was applied in Ethiopia as some aspects got more focus than other ones and would have been therefore regarded/valued more important than other problems. For instance, one parameter was “toe in/toe in” as was “limb abnormal”. Clearly, there is an imbalance regarding the relevance of these two conditions. It was decided to look at both right legs of the individual donkey. This proved to be difficult as donkeys in general are not used to be touched on the hind legs. Quite often this resulted in missing data as it was not possible to touch the donkey without help. Therefore one suggestion is to skip the right hind leg when doing a population-wide assessment also with regard to safety and time issues (PRITCHARD et al., 2005). However, the then obtained results might not accurately display the prevalence of inflammation or sole surface problems because three legs would not be assessed. Considering the extremely high prevalence of lameness reported by BROSTER et al. (2009) in horses and by REIX NEE BROSTER et al. (2014) in draught donkeys lameness should be accurately assessed in order to prevent further suffering.

As the DS team was familiar with using **body mapping** for health check and diagnosis (especially within their normal application of veterinary clinics when treating individual animals) it was decided to use this tool to assess the prevalence of **wounds** (ring finger). Although applicable in an easy way during the actual welfare assessment the body mapping proved to be less suitable when entering and analysing the data. It is complicated, error-prone and not feasible for a population-wide welfare assessment. Although easy to apply in the field the obtained data on wounds has to be categorized in a next step in terms of

location and severity/depth. Within this project this was done for each individual donkey by using a screen of the different body regions. Additionally due to combining wound depth and size a large set of combined variables (e.g. 1A. 1B. 2A. 2B) existed which then had to be assigned to the corresponding body region. Therefore multiple columns for each body region comprising the amounts of a wound were the result of entering the obtained data. This described procedure would occupy much time regarding a population-wide assessment including hundreds of animals. Furthermore – although using a screen – some locations of wounds cannot exactly be assigned to a body region due to the demarcations of the mask. However BURN et al. (2009) state that this is also a problem when using predefined body regions. Nevertheless, it is suggested to switch to a list of predefined body regions possibly combined with a less detailed classification of wounds. For further developments of the protocol wound size could be assessed differently from the definition used in the current study and the results support that most of the wounds detected were of size A. 90.5 % of the donkeys suffering from wounds had had or had a wound of the size of up to two fingers (index and middle finger). A possible solution is to reduce the classifications regarding wound size to two categories (up to two fingers. larger than two fingers). It is also argued that assessing the wound depth gives more information on the welfare state of the donkey. Especially with regard to swellings whose importance with regard to animal welfare was not put into focus during the development of the welfare assessment protocol it is more relevant to detect the severity of a wound than having a detailed classification of size. Related to this aspect the first category “1” which is defined as “hairless, old wounds, scar” mixes up two – per se important – strings of information. There is a difference between a hairless spot and an old wound (white hair) that – especially with regard to the work of an animal welfare charity – has to be considered. Whereas a more or less fresh hairless spot is an acute problem an old wound or a scar might also give information on the effect of intervention strategies of an organisation. For instance, if the result of a donkey welfare assessment in an operating area reveals that there are many healed wounds and no new wounds then this result can be considered as a success in the terms of a working donkey welfare charity. Therefore, category “1” should be redefined possibly also including swellings as part of wounds assessment.

Before analysing the data it was decided to put focus on so-called predisposed areas such as the girth region or the region under the tail which are prone to wounds. In a further development of the protocol it would be a possibility to take account of these predisposed areas only. As wounds are also caused by different factors - e.g. improper harness, hyena bites (conf. BIFFA and WOLDEMESKEL, 2006) - classifications of wounds defined by their causal agent could be a means to detect the fields of improvement and intervention strategies in the operating areas. However, every wound regardless of the region is

important. There is a danger of putting too much focus on these predisposed areas. Therefore it is suggested to further assess wounds with defining predisposed areas and non-predisposed areas to 1. not missing any wounds and 2. still being able to identify prone wound areas and their potential causal agents in order to start interventions.

In accordance to SØRENSEN et al. (2001) some parameters – such as lameness parameters - in the first drafts of the protocol were too detailed for a welfare assessment, lacked the feasibility in the field and were therefore excluded after the pilot testing. For that reason also diseases such as rabies and tetanus (as part of **other signs of injury and disease –** little finger) were excluded although the DS team regard these as highly important due to a perceived high prevalence. It was argued by the author that these diseases would need a detailed diagnosis of an affected donkey (this was supported by other DS team members who mentioned the difficulty of assessing/diagnosing diseases in the field during a welfare assessment). Furthermore it would be of more interest to record the incidence of these diseases (i.e. the rate of new infections in a given population over a specific period of time) rather than the prevalence (i.e. the proportion of infections in a population at a given time) in order to carry out intervention strategies. However, the last row “other signs of diseases” gives the opportunity to state the health status of an animal. It was not used for this purpose though during the assessment which does not mean that these diseases did not occur but they were maybe not detected.

Regarding the layout the **finalised version** which was applied in Ethiopia consisted of **two pages** - usually printed out - on one page. One donkey was assessed per sheet. During the discussions the DS team stated that assessing one donkey with one sheet prevents from forgetting filling out parameters. Regarding the results from the donkey welfare assessment and the inter observer-reliability this assumption cannot be supported as the author and the team members overlooked parameters several times - possibly due to lack of practice/training or time constraints. Furthermore regarding a potential population wide donkey welfare assessment this approach with using one sheet per donkey might not be feasible as the amount of paper needed and carried in the field should not be underestimated. One suggestion would be to re-evaluate specific parameters and assessment methods (e.g. body mapping), develop a protocol which allows to assess more animals on one sheet of paper and test this protocol in the field. This would, however, involve new discussions on the definitions of parameters (e.g. wound size and depth) and training of potential assessors.

Further development of the protocol and of the guidelines would also require **further discussions** in order to detect and define some problem areas more in detail such as mud fever, mange and wounds with regard to other signs of injury and disease (Little finger). Mud fever (Photo 9) for instance was firstly assessed as an abnormal condition as flies infestation

on the donkeys' legs. Only during later assessments it was mentioned by one DS member that this was mud fever; however, this diagnosis may not be correct as the affected legs did not show the typical signs for mud fever (HAMILTON-FLETCHER, 2014). According to HAMILTON-FLETCHER (2014) typical signs include "Small, circular, ulcerated, moist lesions beneath scabs" and "Thick, creamy, white, yellow or greenish discharge (…)". These signs were not detected during the assessment of the working donkeys in Ethiopia. Therefore more precise definitions on this condition as well as its welfare relevance must be discussed in order to make clear statements. Additionally the guidelines and the protocol have to be adjusted as there were ambiguous instructions in the chapter dealing with lameness assessment which most likely lead to ambiguous if not false results. Bearing this in mind as well as the restricted amount of time and animals assessed in Ethiopia the finalised protocol and the guidelines can only be seen as a first step to a population-wide donkey welfare assessment.

6.2 Donkey welfare assessment, IOR and QBA

6.2.1 General aspects

The **donkey welfare assessment** with the extended Hands-on Donkey Welfare assessment protocol was carried out with 107 pack donkeys on markets and mill houses in the area around Debre Zeit. Considering the huge population of working donkeys in Ethiopia this sample size must be considered very small. Additionally the welfare assessment was carried out in a localised area in Ethiopia which may also account for specific results such as high prevalence of wounds (comp. to BIFFA and WOLDEMESKEL, 2006). Due to this restricted application of the extended protocol (area, work type) the results do not allow a comprehensive description of the welfare status of working donkeys neither in Ethiopia nor in other operating areas of the DS.

Inter-observer reliability (IOR) was tested by assessing 32 donkeys using the extended Hands-on Donkey Welfare assessment protocol which was also used for the donkey welfare assessment of 107 donkeys carried out by the author. Similar to this assessment the amount of animals used for testing IOR must be considered as rather low and therefore the results must be interpreted carefully. Furthermore, specific statistical problems were detected when the data obtained from the IOR testing were processed. Firstly, all three observers scored at least once as a constant (i.e. one category as always absent or present). Whereas those records are not considered as a problem for calculating percentage agreement (PA) it can be regarded as a problem when calculating the k-value. Therefore, for further IOR testing it is recommended to have a bigger and more heterogeneous sample (i.e. various degrees of conditions) in order to prevent this problem during the analysis. Secondly, SPSS proved to

be not suitable when calculation weighted k-values. Therefore an online tool had to be used for that purpose.

Secondly, a **more heterogeneous sample** is preferable due to interpretation purposes of the k-value as it can be influenced by other factors. Two factors or paradoxes are described in relation with the k-value: **prevalence index** (PI) and **bias index** (BI). A completely heterogeneous (i.e. balanced. various degrees of conditions) population has a PI of 0 whereas the opposite of 1 describes an unbalanced, therefore homogenous population (BURN et al., 2009). A large PI reduces the k-value whereas this does not occur with a low or zero PI. If an attribute (e.g. a clinical condition. classified as absent or present) is highly prevalent in a given population the PI is equally high as the proportion of positive and negative classifications has an effect on the ratings of two observers. Whenever the prevalence of a positive rating is very high or very low the PI is high as the agreement by chance is relatively high as well. Therefore, the k-value is reduced. Given the same number of agreements the k-value derived from a balanced sample of positive and negative classifications is higher than compared to an unbalanced pattern (e.g. only positive classifications. no negative classifications) (SIM and WRIGHT, 2005). Therefore, a low k-value can either indicate a generally low agreement between two observers or the population was too homogeneous which makes – as already mentioned - agreement by chance very likely.

Bias describes to what extent two observers disagree in their positive or negative answers or if their pattern of disagreement is symmetrical and is expressed by the bias index (BI). Contrary to the PI with a high BI and therefore a large bias the Kappa value is also higher than compared to low or zero BI (SIM and WRIGHT, 2005). The effect of prevalence is greater for a large k-value – in contrast, the bias effect is greater with smaller values of Kappa than larger ones.

PABAK stands for *prevalence adjusted bias-adjusted kappa* and assumes that there are no prevalence or bias effects what is critically considered from several sides (SIM and WRIGHT, 2005, BURN et al., 2009). However, it also suggested to present the PABAK alongside with the Kappa value in order to show the effects of prevalence and bias (SIM and WRIGHT, 2005).

The influence of prevalence on the interpretation of the k-value can be seen in the following examples from this present IOR analysis: from the *k* value of less than -0.01 for “Eyes discharge” no agreement may be assumed (i.e. less agreement than by chance). However when looking at one pairwise comparison two observers had a PA of 94 % but nevertheless a Kappa value of -0.033. The PI is very high (0.94) indicating an extremely homogenous population and therefore possibly influencing the k-value. Thus it is important to present the PI alongside the k-value as well as the PA for interpretation purposes as within this given example the k-value as a standalone result could lead to false conclusions.

According to BYRT et al. (1993) the effects of bias in the present study seem to be relatively small. For instance, the k-value of “Nose discharge” is 0.12, with 57 % PA, a PI of 0.26 and a comparatively high BI of 0.24. However, the influence of BI must be considered low as the k-value is already low. Here, the main conclusion is that agreement between the observers needs to be improved - the BI does not provide additional information. Additionally, as the PI is very high regarding a substantial numbers of parameters the bias effects might be suppressed even more. Therefore, future studies in working donkey populations should focus more on the prevalence effects and the interpretation of the k-values as the PIs are usually high due to high prevalence of various welfare issues (BURN et al., 2009). However, if substantial bias occurs BYRT et al. (1993) suggest to put more emphasis on finding the causes for such bias.

6.2.2 Results of donkey welfare assessment, IOR and QBA

The protocol started with a general part including the information on observer, date, location, work type of donkey and sex. Although age was estimated during the teeth assessment the data obtained is regarded as general information. Therefore, regarding the age distribution most of the assessed animals were older than 10 years which is higher compared to results of other studies (PRITCHARD et al., 2005, MCLEAN et al., 2012). IOR testing shows a poor, moderate and substantial agreement which still indicates the need for training of age assessment to achieve a higher consistency.

The first finger represents the dimension of **behaviour and demeanour**. 54.2 % of donkeys appeared to be alert with 45.8 % being apathetic. PRITCHARD et al. (2005) found 88.5 % alert donkeys compared to 11.5 % apathetic/severely depressed donkeys. In another study 46 donkeys appeared to be rather alert than apathetic (MCLEAN et al., 2012). BURN et al. (2010b) report 19.1 % apathetic donkeys with 10843 working equines. One reason for the comparatively high percentage in this study could be that the animals assessed were apathetic as a result of hard work, overloading and long walking distances as the assessments took place at markets and mill houses. However, the assessment of the general demeanour of donkeys within this project is a snap-shot taken during the “break at work” which therefore could display merely tiredness and not a general apathy. Nevertheless, implications for the welfare of these donkeys are given as these apathetic or tired animals still have to perform. Consequences might impair welfare (e.g. stumbling and falling down) or the human-animal relationship (e.g. enforced driving or beating). The inter-observer reliability of this parameter showed moderate agreement between the three observers whereas the results of BURN et al. (2009) display poor inter-observer reliability (but moderate intra-observer agreement) . The same level of agreement was reached within this project in all pair-wise comparisons. A higher k-value is desirable as the original intention of

assessing an animal as “alert”, “apathetic” or “severely depressed” was to reflect on severe conditions. Considering BURN et al. (2010b) who report associations between apathy and several welfare problems such as low BCS it is relevant to detect the behavioural state accurately. Reasons for this result could be caused by a different understanding of the terms “Alert”, “Apathetic” and “Severely depressed”. For instance, if a tired donkey is resting this behaviour could be misunderstood: resting does not necessarily mean that the donkey is apathetic (see above). Secondly, as the donkeys were assessed successively by each observer (three times all in all) they are likely to show a different behaviour every time, maybe as they felt disturbed by the first assessment and seemed to be alert when being assessed by the second observer (BURN et al., 2009). Within this study the inter-observer reliability testing is therefore confounded with test-retest reliability as alterations during the assessment situations occurred. The original intention of assessing an animal as “alert” or “severely depressed” was – however - to reflect on severe conditions which should not change within five minutes and should also not be influenced by the (gentle) assessment of two people or the surroundings of the assessment. Therefore, it is questionable if the assessors used an interpretation according to the above stated (or the training was inadequate or the definition is not precise enough).

The observer approach showed results similar to those of PRITCHARD et al. (2005) with slightly more donkeys showing avoidance and aggressive behaviour than a friendly response. However, the observer approach proved to be difficult to carry out under standardised circumstances as the donkeys were assessed on the market or in the village around mill houses. The donkeys are usually tied together as shown on Photo 1 (p. 22) which makes it difficult to approach the donkey in a standardised way. The protocol was developed under the light of a potential population-wide assessment preferably carried out in a short amount of time and without disturbing the owner unnecessarily. However, in order to get standardised data, it might be necessary to randomly choose individual donkeys and assess those animals in a surrounding which does not comprise a standardised approach. Regarding the IOR of the observer approach the same holds true as stated above: the agreement was poor when assessing if the donkey does not move or moves head away. However it was moderate when the donkey moves away. It is possible that the donkey showed a different behaviour each time regarding the first two categories. The moderate agreement for “Donkey moves away” suggests that donkeys stick to this behaviour of avoiding the observer each time. Furthermore, the result indicates that this category is more unambiguous compared to the other categories.

Additionally, it was mentioned during the discussions with the DS team that there might a difference between the approach of an unknown person and a familiar person (i.e. the owner) (see also POPESCU and DIUGAN, 2013). It was argued that the familiar owner more

likely evokes a friendly response compared to an unfamiliar observer. However, approach tests are carried out in order to assess if animals perceive humans as positive, negative or neutral stimulus assuming a generalized reaction. Horses, for instance, are considered to display a general response to humans when their behavioural response to an approaching. unfamiliar observer is tested (HAUSBERGER et al., 2008). However, POPESCU and DIUGAN (2013) report different behavioural reaction towards a known and unknown person in working horses in Romania, with more animals showing avoidance/fear towards the unknown observer in the approach test. More than half of the horses showed indifference towards the unknown person. In the present study, with 50 % of the donkeys showing avoidance behaviour during the approach test it is still questionable if humans in general can be regarded as a positive or neutral stimulus. Instead, it might indicate fear of humans and therefore a welfare issue considering the working environments of these pack donkeys (i.e. markets and mill house and surrounded by people). Furthermore, it must be questioned to what extent a horse's reaction can be compared to a donkey's behavioural answer towards an observer approach. Additionally, an approach test might not be accurate to test the quality of the human-animal bond (the animal and its owner). Especially, a "bond of interdependence" between animal and owner is regarded as highly important in regard to potential welfare interventions as a greater amount of motivation for change of practice and improvement is considered to exist whenever the relationship goes beyond a mere economic level (i.e. earning money) (PRITCHARD et al., 2005).

The chin contact and ear test were slightly altered by changing and extending the categories as mentioned above (p.53). The assessments within this project revealed higher scores than similar earlier studies: while in the present study 40 % and 29 % of the donkeys showed avoidance behaviour during the chin contact and ear contact test respectively PRITCHARD et al. (2005) report 18.5 % of 2596 assessed donkeys showing avoidance during the chin contact and MCLEAN et al. (2012) found avoidance in 13% (7 out of 53 animals) in an ear contact test. Furthermore, a large percentage (29 % and 48 %) could not be assessed due to preceding avoidance behaviour (i.e. moving away).

According to the Ethiopian DS chief of staff donkeys generally avoid being touched at their ears but some severely depressed animals might not react to this contact test at all. However, in the present study no severely depressed animals were detected because the donkeys observed still reacted towards their environment which is not the case according to the definition of severe depression. Therefore, this assumption must be examined in a further study.

The **body condition** scoring represents the index finger. The prevalence of donkeys with a low body condition (score 1 and 2) is with more than 98 % very high. which is in accordance with other studies (MCLEAN et al., 2012, BURN et al., 2010a, MEKURIA and ABEBE, 2010) –

however, the results of this study display a high prevalence of a moderate body condition (score 2) whereas other studies report a higher percentage of donkeys with poor body condition (score 1) (PRITCHARD et al., 2005). BURN et al. (2010a) argue that the BCS is a useful indicator for other welfare problems such as wounds or unresponsiveness towards handling/environment. Contrary to BURN et al. (2009) the k-value shows variable agreement between the observers (poor, substantial and moderate agreement). However, as one of the DS team members assigned half scores in the first round the IOR testing should be repeated in order to confirm this result in order to obtain a more reliable result.

Lameness is considered as the middle finger of the Hands-on Donkey Welfare assessment protocol. A high percentage of the 107 animals were assessed as (potentially) lame. Other studies have equally shown that the high prevalence of lameness is a serious welfare issue in working equines (BROSTER et al., 2009, PRITCHARD et al., 2005) but it needs to be taken into account that different studies use different ways of assessing lameness. Such different definitions of lameness and scoring schemes may result in different prevalence rates.

In the present study the approach was chosen to assign the animals to the categories “not lame”, “potentially lame” and “obviously lame” as the DS team argued that donkeys do not easily show lameness or gait abnormality. Therefore, any problem perceived such a hoof abnormality could be seen as a potential source of pain and lameness. This might lead to a higher prevalence of donkeys having lameness-associated problems (e.g. hoof too long).

Weight shifting was not observed but this may be expected since - according to the DS team - this behaviour is mainly shown by packed donkeys. The vast majority of the donkeys were not packed when the assessment took place. As most of the donkeys were not used to be touched on the hind leg the results primarily display abnormalities of the right front leg. If the records are complete (i.e. both legs assessed) it is however not possible to trace back if the front leg or the hind leg is affected. Therefore, no suggestion can be made from this data to just assess the front right leg. However, as mentioned before, for health and safety as well as feasibility reasons assessment of the front legs only should be preferred (PRITCHARD et al., 2005) although including the high risk of underestimating the lameness prevalence in working donkeys (REIX NEE BROSTER et al., 2014). The IOR results of the lameness assessment cannot be discussed without bearing in mind the restrictions of the guidelines and the protocol which are partly inconsistent (e.g. weight bearing). As mentioned on p. 28 the underlying concept of the protocol is to state whether some problem is absent (0) or present (1) which was – however – inconsistently described. As the IOR for lameness showed poor agreement in all parameters, more detailed, clear and non-misleading definitions and training are necessary.

Wounds (ring finger) are not only in Ethiopia but anywhere where working donkeys are used a welfare issue (PRITCHARD et al., 2005). In this study the amount of donkeys with one or

more wounds is high with almost 88 %. This could be due to the region and the sample size. The sample size with 107 donkeys is much smaller compared to PRITCHARD et al. (2005) with n=2596 donkeys. PRITCHARD et al. (2005) report breast/shoulder (11.5 %), withers (10.2 %) and girth (18.3 %) as well as hind quarters (12.3%) as predominant areas for wounds in working donkeys. Taking the two different classifications of body parts in the two protocols used into account (Working Equine Welfare Assessment (PRITCHARD and WHAY, not stated) vs. Hands-on Donkey Welfare assessment protocol) the results still show a difference of prevalence of back wounds. Prevalence of back wounds in PRITCHARD et al. (2005) was 29.8 % (calculated by the author combining prevalence scores of withers. spine and hind quarters) with displaying only deeper lesions (full thickness skin or deeper lesions) not including superficial or healed wounds. In the present study – when only analysing wounds with wound depth 2 and 3 (no score 4 was obtained) in n=107 donkeys - the value is 19.6 %. Another study carried out in three regions in Ethiopia showed that 58 % of 922 assessed donkeys were suffering from back sores (TESFAYE and MARTIN CURRAN, 2005). BIFFA and WOLDEMESKEL (2006) report a prevalence of 79.4 % of external injuries – compared to 88 % of the donkeys having wounds in the present study - stating also that improper harness and saddle design are the most important reason for injuries. The authors also state that donkeys are having more wounds than horses contrary to common assumptions that donkeys are more tolerant to the working conditions given in Ethiopia. Although widely used little care is taken of the working donkeys according to the authors. The findings regarding wounds support the results of the present study including hyena bites as a source of concern (see also TESFAYE and MARTIN CURRAN, 2005).

The IOR for wounds was assessed by recording if the three observers saw a wound in a specific body area (size. depth or amount of wounds was disregarded). The IOR ranges from poor (wounds on the legs) to substantial (back wounds). One possible explanation for this wide range is the circumstances of the welfare assessment carried out on the markets and villages. As the donkeys were kept as shown on Photo 1 wounds can be easily overseen as the donkeys stand close together and can sometimes not be assessed properly. As stated above this suggests to carry out a welfare assessment in a more structured way by randomly selecting the animals and checking their welfare status in a situation where the animal is better accessible.

Regarding the little finger - **other signs of injury and disease** - the prevalence of ectoparasites is high compared to other studies (PRITCHARD et al., 2005). The infestation with Gasterophilus is most common in the region around Debre Zeit. The agreement between the different observers was moderate (in two pair-wise comparisons) and substantial still indicating a need of improvement of definitions and training. Compared to this value the assessment of coat condition proved to be highly inconsistent. This could be due to lack of

training but also the occurrence of dirty animals which could make it difficult to score the coat condition correctly. Similarily to BURN et al. (2009) the observers could also have experienced difficulties to distinguish the cut-off points between the three scores “excellent”. “average” and “poor”.

According to the results in the present study heat stress was not a problem which is most likely due to the weather conditions of that time of the year (August-September. rainy season). In order to detect problems with assessing heat stress it is advisable to test the protocol in a different season or different location (PRITCHARD et al., 2006). The IOR in the present study revealed a PA of 98 %, with a k-value of 1 (pair-wise comparison Observer 1 and 3) despite a high PI of 0.94. BURN et al. (2009) report ambiguous IOR scores in their study, also showing a high PI which might have influenced the magnitude of the k-values.

Generally speaking the amount of animals assessed with the developed protocol as well as the restriction of one region in Ethiopia and the work type of the animals (only pack donkeys) are as already mentioned major impact factors. A larger number of working animals including different work types would have conveyed more valid results which also holds true for the IOR. However - due to time restrictions - this was not feasible but must be considered in further studies regarding the Hands-on Donkey Welfare assessment protocol before it is applied and used as a tool for population-wide donkey welfare assessments.

The use of **Qualitative Behaviour Assessment (QBA**) aimed at possibly developing an additional measure for the first finger behaviour and demeanour. It was carried out for the first time with donkeys in order to apply a very practical, not time-consuming and holistic method. In the beginning of this project the intention of applying QBA with working donkeys was more to put the focus on the “whole animal” aspect and to try out a new, different as purely qualitative welfare assessment method.

The first component – described as “General mental state” and including terms such as “happy”, “active” or “curious” on the positive side and “depressed” and “exhausted” on the negative side – showed a good overall consistency between the three different observers. The dimension or terms describing the emotional and mental state has been reported in several studies dealing with QBA in other animals: dairy cows (ROUSING and WEMELSFELDER, 2006), sheep (WICKHAM et al., 2012), dogs (WALKER et al., 2010), horses and ponies (NAPOLITANO et al., 2008). However, PCA at the single observer level revealed that the observers had a general common understanding of some specific terms (such as “happy” and “friendly”) whereas this did not hold true for other terms (“relaxed”). One possible explanation for the lack of common understanding of e.g. “relaxed” could be that the donkeys observed reacted hardly to their environment (market place. many people) which could be interpreted as a “relaxed way of coping with surroundings”. However, this could be falsely interpreted as these animals might also be exhausted and tired and therefore they

were not able to react to the environment. This cannot be regarded as a relaxed attitude or behaviour. Therefore more in-depth discussions about the descriptive terms and their different levels and specifications may have led to clearer results.

Also the scores of “General mental state” for each individual donkey (p. 52) revealed that the observers assessed some donkeys differently. Here, one possible explanation might lie in the application itself: As the assessment was carried out consecutively (around 8 minutes between two observations), the behavioural expression of the donkeys might have changed in the course of the observations by the three different observers. Further studies with more animals and observers are needed. Furthermore there remains the question of and how far QBA provides additional information on working donkeys’ welfare. The behavioural parameter “General demeanour” - which is originally part of the Working Equine Welfare Assessment protocol (PRITCHARD and WHAY, not stated, PRITCHARD et al., 2005) - already comprises a qualitative dimension. However, the three variables “Alert”, “Apathetic” and “Severely depressed” might not be sufficient enough to cover the body language of working donkeys. Here, QBA has the potential to provide a wider and deeper insight and information. Nevertheless the question to what extent QBA correlates with other behavioural and health parameters cannot be sufficiently answered within this thesis as the results obtained from the first-time application of QBA are not reliable enough for that purpose.

It must be questioned if nine terms are sufficient to assess donkeys’ behaviour comparing other animal welfare assessment protocols which include QBA with a fixed list of 20 terms (WEMELSFELDER et al., 2009a, WEMELSFELDER et al., 2009b). Within this project limitations were given in terms of number of assessors and the novelty and unfamiliarity of this qualitative approach. Therefore, the given results can just be interpreted as a first step of QBA within donkeys. Additional research should be carried out focusing on how observers with different professional/social (e.g. owners. veterinarians. scientists) and geographical (e.g. Europe. Asia. Africa) backgrounds score working donkeys. Although the last round of QBA with the grazing herd of donkeys does not provide statistically analysed results the discussions after the observation showed that the different observers (three Ethiopians. one Austrian) have different experiences of a friendly, curious donkey that is not feeling discomfort. This can be seen in the raw data (i.e. ticks on the scale). It was mentioned from one team member that the Ethiopian DS team has not seen donkeys yet kept as merely companion animals. Therefore, they might have considered this grazing herd of working donkeys as not feeling discomfort and friendly as these animals do not show these typical signs of discomfort that were mentioned such as being tethered at one place, not being able to defend itself from flies and having itchy spots. However, it can be argued that these descriptions lack the subjective level (i.e. How does the donkey react to the flies? How does it feel?) as a tethered donkey can still feel happy although flies are annoying it. External

factors should not be the major factor for the qualitative assessment of working donkeys' behaviour but how the donkeys feel and react towards these external factors. Therefore it must be questioned if the underlying concept of QBA was thoroughly explained and applied. Another reason could be the lack of experience of the author scoring the expressive quality of working donkey behaviour. A third reason might be that the terms and their definitions are not concise enough: a "minimally friendly" donkey is not necessarily unfriendly but can also show "indifference". Therefore this suggests that – as mentioned above – the amount of terms is not sufficient. Terms expressing the opposite such as "aggressive/hostile" in the case of "friendly" would most likely reveal more information on how the observers perceive and assess the behavioural style of donkeys. Furthermore there is a need for discussion in what way QBA can be implemented within a population-wide assessment. One suggestion could be to score donkeys along a row of boxes (e.g. four or six) instead of a scale although this approach must be tested thoroughly as it would include a new approach to QBA not including the relevant principle of the Visual Analogue Scale. Any implementation of QBA can only take place if further studies dealing with above mentioned aspects are carried out beforehand.

7 Conclusion

The developed protocol is based on the qualitative Hands-on Donkey welfare assessment Protocol and was extended and deepened by collating parameters of already existing protocols as well as a literature review. The parameters were chosen according to the five fingers of the original Hands-on Donkey Welfare version. Furthermore, the experiences and knowledge of the DS team and the pilot testing in Ethiopia were an important part of the development process. Carrying out the welfare assessment within a standardised approach as well as the adaptation of QBA on working donkeys is a first step towards a population-wide donkey welfare assessment with in the DS. The results of the donkey welfare assessment support previous studies and show that wounds, low BCS and limb/hoof problems are highly prevalent. QBA was carried out for the first time with working donkeys and was adapted to assess the behaviour of these animals in a qualitative way, identifying two main components which were interpreted as “General mental state” and “Responsiveness towards social stimuli”. Especially the first component describing mental state shows a good overall consistency between the three different observers. Observer-wise analyses of QBA also reveal that there is a common understanding of some specific terms (e.g. active) whereas other terms such as “feeling discomfort” do have different meanings for the individual observers. These first results show that it was possible to carry out QBA in donkeys with – regarding this method – inexperienced observers although there is room for improvement regarding explaining and applying QBA with working donkeys. They indicate a need for further research of QBA in these animals including more observers and animals, with possibly generating more descriptive terms of the donkeys’ behaviour. Furthermore, there is a need of further studies in this area in order to find out its suitability in the context of a broader-scale donkey welfare assessment. Additionally, IOR testing revealed an inconsistent assessment regarding more than half of the parameters in the protocol which calls for improvement on several levels: Firstly, more in-depth discussions on the definitions and training are highly advisable. As the training might have been not sufficient enough in this round of IOR testing, more in-depth training might resolve problems with inconsistent rating, especially since some parameters and assessment methods were new to the observers. Age estimation by assessing the teeth is commonly applied within the everyday work of DS (e.g. diagnosing health problems during a clinic day) and here the assessment could be still improved. Secondly, the sample size for IOR testing must be larger and more diverse in order to minimize potential PI effects as well as computing problems. One solution might be to use classroom tests with photos showing different conditions. However, there are limitations to this approach as for instance behavioural parameters might not be that clear to assess. The protocol was used to assess pack donkeys in one region in Ethiopia - further

development of the protocol and extensive testing with more animals and in different regions is therefore necessary for a future population-wide usage of the developed protocol.

8 Abstract

Donkeys play a crucial role as a work force in many regions worldwide. These animals may suffer from various welfare problems such as lameness and wounds.

The animal welfare charity *The Donkey Sanctuary* (DS) works towards an improvement of this situation. The animal welfare charity *The Donkey Sanctuary* (DS) operates overseas to improve the welfare of working donkeys. To monitor the welfare situation and to evaluate the impact of interventions aiming at welfare improvement, a standardised, animal-based welfare assessment is necessary. So far, this assessment has been carried out subjectively by using the Hands-on Donkey Welfare assessment protocol. In the present study, parameters of existing protocols were collated and discussed with the DS team in order to develop a more elaborate protocol using a standardised approach. For the first time in donkey welfare assessment, the protocol also included qualitative behaviour assessment (QBA). The final version was tested with 107 pack donkeys in Ethiopia.

The results regarding the behavioural assessment show that 45.8 % of the donkeys were apathetic. The reaction to the observer approach supports this result: 44.3 % of the animals did not move whereas 31.1 % moved away. QBA revealed three main components: the first main component – described as “General mental state” – showed a good overall consistency between the three different observers.

The vast majority (98.5 %) of the donkeys assessed had a low body condition score. The prevalence of wounds and lesions was high with 88 % of the animals having one or more wounds including healed lesions.

Additionally, inter-observer reliability was tested between three observers by assessing 32 donkeys. It showed an inconsistent rating regarding more than half of chosen parameters. Thus, more in-depth discussions on the definitions and training are highly advisable.

The developed welfare assessment protocol including objective parameters and the first adaptation of QBA with working donkeys is one step to a population wide assessment of donkeys to be used for the field work of the DS. However, further developments and additional testing in different regions and with more animals are necessary.

9 Zusammenfassung

Die Bedeutung von Arbeitseseln ist in vielen Regionen der Erde sehr groß. Vielfach leiden diese Tiere unter Beeinträchtigungen des Wohlergehens wie z.B. Wunden oder Lahmheit.

Die Tierschutzorganisation *The Donkey Sanctuary* (DS) arbeitet hier an einer Verbesserung dieser Situation. Um den Effekt der Arbeit zu evaluieren, ist eine standardisierte Beurteilung des Wohlergehens notwendig. Bis jetzt beruhte diese rein auf einer subjektiven Einschätzung mithilfe des Hands-on Donkey Welfare Assessment Protokolls.

In diesem Projekt wurden für dieses Protokoll Parameter bestimmt, um eine tiefergehende. standardisierte Beurteilung zu ermöglichen. Die Parameter wurden einerseits aus bereits bestehenden Protokollen übernommen, andererseits mit DS-Mitarbeitern aus UK und Äthiopien diskutiert und festgelegt. Ein weiterer Aspekt zur Verhaltensbeurteilung war das Qualitative Behaviour Assessment (QBA.).

Die endgültige Version wurde an 107 Packeseln in Äthiopien getestet. Bei der Verhaltensbeurteilung zeigten sich 45.8% der Esel apathisch. Dem entsprach die Reaktion auf die Annäherung der Beurteilerin: 44.3% der Tiere bewegten sich nicht, während 31.1% sich wegbewegten. Bei QBA wurden drei Hauptkomponenten extrahiert. Die erste Hauptkomponente „Allgemeine Befindlichkeit“ erwies sich als konsistent über alle drei BeurteilerInnen hinweg. Die Ergebnisse zeigen des Weiteren, dass 98.5% der Esel unterkonditioniert sowie Wunden oder Läsionen hoch prävalent waren (88 % der Tiere mit mindestens einer Wunde, weißer Haarstelle oder Narbe). Zudem wurde die Inter-Observer Reliabilität zwischen drei Beobachtern anhand von 32 Eseln getestet. Bei mehr als der Hälfte der Parameter zeigte sich keine zufriedenstellende Übereinstimmung. Tiefergehende Diskussionen rund um die Definitionen und mehr Training sind daher ratsam.

Durch die Beurteilung anhand objektiver Indikatoren inklusive der erstmaligen Adaptierung von QBA auf Arbeitsesel ist ein erster Schritt zur Wohlergehensbeurteilung von Eseln auf Populationsebene für die Anwendung durch die DS getan. Es bedarf aber noch einer Weiterentwicklung und zusätzlicher Anwendungen in anderen Regionen und an mehr Tieren.

10 Literature

Ashley, F. H., Waterman-Pearson, A. E. & Whay, H. R. (2005). Behavioural assessment of pain in horses and donkeys: application to clinical practice and future studies. *Equine Veterinary Journal* [Online], 37 (6), 565-575. Available: http://dx.doi.org/10.2746/042516405775314826 [Accessed 18.07.2013].

Ashley, F. H., Waterman-Pearson, A. E. & Whay, H. R. Development of an ethogram to record potential behavioural indicators of pain in working donkeys. *In:* PEARSON, R. A., MUIR, C. J. & FARROW, M., eds. Proceedings of the Fifth International Colloquium on Working Equines. The Future for Working Equines, 30th October - 2nd November 2006 Addis Ababa, Ethiopia. 15 - 24.

Ayele, G., Feseha, G., Bojia, E. & Joe, A. (2006). Prevalence of gastro-intestinal parasites of donkeys in Dugda Bora District, Ethiopia. *Livestock Research for Rural Development* [Online], 18 (10), Article #136. Available: http://dx.doi.org/http://www.lrrd.org/lrrd18/10/ayel18136.htm [Accessed 06.05.2013].

Bekele, Z. (2004). Traditional sayings about donkeys in Ethiopia. *In:* STARKEY, P. & FIELDING, D. (eds.) *Donkeys, people and development. A resource book of the Animal Traction Network for Eastern and Southern Africa (ATNESA). Web version.* Wageningen, The Netherlands: ACP-EU Technical Centre for Agricultural and Rural Cooperation (CTA), http://www.atnesa.org/donkeys/donkeys-zelealam-sayings-ET.pdf [Accessed 19.07.2013].

Biffa, D. & Woldemeskel, M. (2006). Causes and Factors Associated With Occurrence of External Injuries in Working Equines in Ethiopia. *International Journal of Applied Research in Veterinary Medicine* [Online], 4 (1), 1-7. Available: http://dx.doi.org/http://www.jarvm.com/articles/Vol4Iss1/Vol4Iss1BiffaV4N1pp1-7.pdf [Accessed 06.05.2013].

Blakeway, S. (2013). *RE: Hands-on Donkey Welfare Assessment.* Personal communication with ZEDLACHER, H.

Blench, R. M. (2000). *A history of donkeys, wild asses and mules in Africa,* London: University College Press. http://www.rogerblench.info/Ethnoscience/Animals/Livestock/Donkeys%20in%20Africa.pdf [Accessed 15.07.2013].

Blench, R. M. (2004). The history and spread of donkeys in Africa. *In:* STARKEY, P. & FIELDING, D. (eds.) *Donkeys, people and development. A resource book of the Animal Traction Network for Eastern and Southern Africa (ATNESA). Web version.* Wageningen, The Netherlands: ACP-EU Technical Centre for Agricultural and Rural Cooperation (CTA), http://www.atnesa.org/donkeys/donkeys-blench-history.pdf [Accessed 15.07.2013].

Broom, D. M. (1996). Animal Welfare Defined in Terms of Attempts to Cope with the Environment. *Acta Agriculturae Scandinavia. Section A. Animal Science* Supplementum 27**,** 22 - 28.

Broom, D. M. & Fraser, D. (2007). *Domestic animal behaviour and welfare.* Wallingford, Oxfordshire: CABI.

Broster, C. E., Burn, C. C., Barr, A. R. S. & Whay, H. R. (2009). The range and prevalence of pathological abnormalities associated with lameness in working horses from developing countries. *Equine Veterinary Journal* [Online], 41 (5), 474-481. Available: http://dx.doi.org/10.2746/042516409x373907 [Accessed 25.04.2013].

Bryceson, D. F. (2010). Modernizing rural livelihoods and transport in Africa: Directions and dilemmas. *In:* THE BROOKE (ed.) *The 6th International Colloquium on Working Equids: Learning from Others. Proceedings of an International Colloquium held at the India Habitat Centre, New Dehli, India, 29th November to 2nd December 2010.* The Brooke, 30 Farringdon Street, London, EC4A 4HH, UK, 76 - 79,

http://www.thebrooke.org/__data/assets/pdf_file/0007/49381/Theme3-Livelihoods-rural-transport-and-the-global-context-45-64.pdf [Accessed 26.07.2013].

Burden, F. A., Toit, N., Hernandez-Gil, M., Prado-Ortiz, O. & Trawford, A. (2010). Selected health and management issues facing working donkeys presented for veterinary treatment in rural Mexico: some possible risk factors and potential intervention strategies. *Tropical animal health and production* [Online], 42 (4), 597-605. Available: http://dx.doi.org/10.1007/s11250-009-9462-0 [Accessed 20.11.2013].

Burn, C. C., Dennison, T. L. & Whay, H. R. (2010a). Environmental and demographic risk factors for poor welfare in working horses, donkeys and mules in developing countries. *Veterinary journal* [Online], 186 (3), 385-92. Available: http://dx.doi.org/10.1016/j.tvjl.2009.09.016 [Accessed 01.05.2013].

Burn, C. C., Dennison, T. L. & Whay, H. R. (2010b). Relationships between behaviour and health in working horses, donkeys, and mules in developing countries. *Applied Animal Behaviour Science* [Online], 126 (3-4), 109-118. Available: http://dx.doi.org/10.1016/j.applanim.2010.06.007 [Accessed 25.04.2013].

Burn, C. C., Pritchard, J. C., Farajat, M., Twaissi, A. a. M. & Whay, H. R. (2008). Risk factors for strap-related lesions in working donkeys at the World Heritage Site of Petra in Jordan. *Veterinary journal* [Online], 178 (2), 263-71. Available: http://dx.doi.org/10.1016/j.tvjl.2007.07.014 [Accessed 18.06.2013].

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

Byrt, T., Bishop, J. & Carlin, J. B. (1993). Bias, prevalence and kappa. *Journal of Clinical Epidemiology* [Online], 46 (5), 423-429. Available: http://dx.doi.org/10.1016/0895-4356(93)90018-V [Accessed 10.1.2014].

Cannon, R. C. & Perrott, J. C. G. (not stated). *Welfare Assessment in Working Equids.* University of Bristol.

Cyril Roy, C. Influence of dental health on some welfare indicators of working. *In:* PEARSON, R. A., MUIR, C. J. & FARROW, M., eds. Fifth International Colloquium on Working Equines. The Future for Working Equines., 30 October-2 November 2006 Addis Ababa, Ethiopia. 216 - 219.

Dewhurst, K. & Henstridge, C. (not stated). *The Development of a Checklist to Assess the Health and Welfare of Equids.* University of Bristol.

du Toit, N., Burden, F. A. & Dixon, P. M. (2008). Clinical dental findings in 203 working donkeys in Mexico. *Veterinary journal* [Online], 178 (3), 380-6. Available: http://dx.doi.org/10.1016/j.tvjl.2008.09.013 [Accessed 18.04.2014].

du Toit, N. & Dixon, P. M. (2012). Common dental disorders in the donkey. *Equine Veterinary Education* [Online], 24 (1), 45-51. Available: http://dx.doi.org/10.1111/j.2042-3292.2011.00236.x [Accessed 18.04.2014].

Duncan, I. J. H. (1996). Animal Welfare defined in Terms of Feelings. *Acta Agriculturae Scandinavia. Section A. Animal Science,* Supplement 27**,** 29 - 35.

Duncan, I. J. H. & Petherick, J. C. (1991). The implications of cognitive processes for animal welfare . I J Duncan and J C Petherick J ANIM SCI 1991 , 69 : 5017-5022 . The online version of this article , along with updated information and services , is located on the World Wide Web at : THE IMPLI. 5017-5022. http://www.ncbi.nlm.nih.gov/pubmed/1808195 [Accessed 14.05.2013].

Edwards, S. (2007). Experimental welfare assessment and on-farm application. *Animal Welfare,* 16**,** 111-115.

FAO/The Brooke (2011). The role, impact and welfare of working (traction and transport) animals. Electronical consultation. http://www.fao.org/fileadmin/user_upload/animalwelfare/FAO-

The%20Brooke%20working%20animals%20e-consultation%20report.pdf [Accessed 16.07.2013].

FAOSTAT (2013). *Donkey Population 2011*. http://faostat.fao.org/ [Accessed 16.07.2013].

Farm Animal Welfare Council (2013). *Farm Animal Welfare Council. Five Freedoms* [Online]. Available: http://www.fawc.org.uk/freedoms.htm [Accessed 10.06. 2013].

Fernando, P. & Starkey, P. (2004). Donkeys and development: socio-economic aspects of donkey use in Africa. *In:* STARKEY, P. & FIELDING, D. (eds.) *Donkeys, people and development. A resource book of the Animal Traction Network for Eastern and Southern Africa (ATNESA). Web version.* Wageningen, The Netherlands: ACP-EU Technical Centre for Agricultural and Rural Cooperation (CTA), http://www.atnesa.org/donkeys/donkeys-fernando-socioeconomic.pdf [Accessed 16.07.2013].

Fleming, P. A., Paisley, C. L., Barnes, A. L. & Wemelsfelder, F. (2013). Application of Qualitative Behavioural Assessment to horses during an endurance ride. *Applied Animal Behaviour Science* [Online], 144 (1-2), 80-88. Available: http://dx.doi.org/10.1016/j.applanim.2012.12.001 [Accessed 28.05.2013].

Forkman, B. & Keeling, L. (eds.) (2009). *Assessment of Animal Welfare Measures for Dairy Cattle, Beef Bulls and Veal Calves. Welfare Quality Reports No. 11,* Cardiff: Cardiff University.

Fraser, D. (2008). Understanding animal welfare. *Acta Veterinaria Scandinavica* [Online], 50 (Suppl 1), S1. Available: http://dx.doi.org/10.1186/1751-0147-50-s1-s1 [Accessed 17.06.2013].

Getachew, M., Feseha, G., Trawford, A. & Reid, S. W. J. (2008). A survey of seasonal patterns in strongyle faecal worm egg counts of working equids of the central midlands and lowlands, Ethiopia. *Tropical animal health and production* [Online], 40 (8), 637-642. Available: http://dx.doi.org/10.1007/s11250-008-9142-5 [Accessed 17.07.2013].

Hamilton-Fletcher, R. (2014). *Mud fever: signs, treatment and prevention* [Online]. Available: http://www.horseandhound.co.uk/horse-care/horse-care-tips/mud-fever-signs-treatment-and-prevention/ [Accessed 28.02.2014.

Hausberger, M., Roche, H., Henry, S. & Visser, E. K. (2008). A review of the human–horse relationship. *Applied Animal Behaviour Science* [Online], 109 (1), 1-24. Available: http://dx.doi.org/10.1016/j.applanim.2007.04.015 [Accessed 25.02.2014].

Hemsworth, P. H. (2003). Human-animal interactions in livestock production. *Applied Animal Behaviour Science* [Online], 81, 158 - 198. Available: http://dx.doi.org/10.1016/S0168-1591(02)00280-0 [Accessed 11.02.2014].

Johnsen, P. F., Johannesson, T. & Sandøe, P. (2001). Assessment of Farm Animal Welfare at Herd Level: Many Goals, Many Methods. *Acta Agriculturae Scandinavica, Section A - Animal Science* [Online], 51 (sup030), 26-33. Available: http://dx.doi.org/10.1080/090647001316923027 [Accessed 17.06.2013].

Jones, P. A. (2012). Personal communication with ZEDLACHER, H.

Kay, G. & Knottenbelt, D. C. (2007). Tetanus in equids: A report of 56 cases. *Equine Veterinary Education* [Online], 19 (2), 107-112. Available: http://dx.doi.org/10.2746/095777307x181320 [Accessed 20.11.2013].

Keeling, L. J. (2005). Healthy and Happy: Animal Welfare as an Integral Part of Sustainable Agriculture. *AMBIO: A Journal of the Human Environment* [Online], 34 (4), 316-319. Available: http://dx.doi.org/10.1579/0044-7447-34.4.316 [Accessed 19.06.2013].

Keeling, L. J., Rushen, J. & Duncan, I. J. H. (2011). Understanding animal welfare. *In:* APPLEBY, M., MENCH, J., OLSSON, I. & HUGHES, B. (eds.) *Animal Welfare.* 2. ed. Wallingford, Oxon [u.a.]: CAB International, 13 - 26.

Knierim, U. & 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.

Leeb, C., Henstridge, C., Dewhurst, K. & 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.

Lund, V. (2002). *Ethics and Animal Welfare in Organic Animal Husbandry. An interdisciplinary approach.* PhD thesis, Swedish University of Agricultural Sciences.

McLean, A. K., Heleski, C. R., Yokoyama, M. T., Wang, W., Doumbia, A. & Dembele, B. (2012). Improving working donkey (Equus asinus) welfare and management in Mali, West Africa. *Journal of Veterinary Behavior: Clinical Applications and Research* [Online], 7 (3), 123-134. Available: http://dx.doi.org/10.1016/j.jveb.2011.10.004 [Accessed 21.06.2013].

Mekuria, S. & Abebe, R. (2010). Observation on major welfare problems of equine in Meskan District, Southern Ethiopia. *Livestock Research for Rural Development* [Online], 22 (3), 1-10. Available: http://dx.doi.org/http://www.lrrd.org/lrrd22/3/meku22048.htm [Accessed 05.05.2013].

Minero, M., Tosi, M. V., Canali, E. & Wemelsfelder, F. (2009). Quantitative and qualitative assessment of the response of foals to the presence of an unfamiliar human. *Applied Animal Behaviour Science* [Online], 116 (1), 74 - 81. Available: http://dx.doi.org/10.1016/j.applanim.2008.07.001 [Accessed 06.11.2013].

Morgan, R. (2007). The epidemiology of lameness in working donkeys in Addis Ababa and the central Oromia region of Ethiopia: a comparative study of urban and rural donkey populations. *In:* PEARSON, R. A., MUIR, C. J. & FARROW, M., eds. Fifth International Colloquium on Working Equines. The Future for Working Equines., 30 October-2 November 2006 Addis Ababa, Ethiopia. 99 - 106

Mullan, S., Edwards, S. A., Butterworth, A., Whay, H. R. & Main, D. C. (2011). Inter-observer reliability testing of pig welfare outcome measures proposed for inclusion within farm assurance schemes. *Veterinary Journal* [Online], 190 (2), e100-9. Available: http://dx.doi.org/10.1016/j.tvjl.2011.01.012 [Accessed 31.01.2014].

Murray, L. M. A., Byrne, K. & D'Eath, R. B. (2013). Pair-bonding and companion recognition in domestic donkeys, Equus asinus. *Applied Animal Behaviour Science* [Online], 143 (1), 67-74. Available: http://dx.doi.org/10.1016/j.applanim.2012.11.005 [Accessed 16.04.2014].

Napolitano, F., De Rosa, G., Braghieri, A., Grasso, F., Bordi, A. & Wemelsfelder, F. (2008). The qualitative assessment of responsiveness to environmental challenge in horses and ponies. *Applied Animal Behaviour Science* [Online], 109 (2-4), 342-354. Available: http://dx.doi.org/10.1016/j.applanim.2007.03.009 [Accessed 06.11.2013].

Olmos, G., Burden, F. & Gregory, N. G. (2010). An innovative approach for better understanding the signs of pain on donkeys: the associations of pain-related pathology with clinical and behavioural indicators. *In:* THE BROOKE (ed.) *The 6th International Colloquium on Working Equids: Learning from Others. Proceedings of an International Colloquium held at the India Habitat Centre, New Dehli, India, 29th November to 2nd December 2010.* The Brooke, 30 Farringdon Street, London, EC4A 4HH, UK, http://www.thebrooke.org/__data/assets/pdf_file/0005/49379/Theme1-Effective-Project-planning-monitoring-and-evaluation-1-39.pdf [Accessed 29.07.2013].

Pearson, R. A., Nengomasha, E. & Krecek, R. (1999). The challenges in using donkeys for work in Africa. *In:* STARKEY, P. & KAUMBUTHO, P. (eds.) *Meeting the challenges of animal traction. A resource book of the Animal Traction Network for Eastern and Southern Africa (ATNESA), Harare, Zimbabwe.* London: Intermediate Technology Publications, 190 - 198, http://www.atnesa.org/challenges/challenges-pearson-donkeys.pdf [Accessed 17.07.2013].

Popescu, S. & Diugan, E.-A. (2013). The Relationship Between Behavioral and Other Welfare Indicators of Working Horses. *Journal of Equine Veterinary Science* [Online], 33 (1), 1-12. Available: http://dx.doi.org/10.1016/j.jevs.2012.04.001 [Accessed 11.07.2013].

Powell, K. (2004). Donkeys and their importance to people. *Work World wide Articles* [Online]. http://www4.worldbank.org/afr/ssatp/Resources/HTML/Gender-RG/Source%20%20documents/Issue%20and%20Strategy%20Papers/G&T%20Rationale/ISGT20%20Value%20of%20Donkeys%20in%20Ethiopia.pdf [Accessed 18.07.2013].

Pritchard, J. & Whay, H. R. (not stated). Guidance Notes to Accompany Working Equine Welfare. University of Bristol.

Pritchard, J. C., Barr, A. R. S. & Whay, H. R. (2006). Validity of a behavioural measure of heat stress and a skin tent test for dehydration in working horses and donkeys. *Equine veterinary journal* [Online], 38 (5), 433-8. Available: http://dx.doi.org/10.2746/042516406778400646 [Accessed 25.04.2013].

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

Pritchard, J. C., Burn, C. C., Barr, A. R. S. & Whay, H. R. (2008). Validity of indicators of dehydration in working horses: A longitudinal study of changes in skin tent duration , mucous membrane dryness and drinking behaviour. 1-7. Available: http://dx.doi.org/10.2746/042516408x297462 [Accessed 25.04.2013].

Pritchard, J. C., Lindberg, A. C., Main, D. C. J. & Whay, H. R. (2005). Assessment of the welfare of working horses, mules and donkeys, using health and behaviour parameters. *Preventive Veterinary Medicine* [Online], 69 (3-4), 265 - 83. Available: http://dx.doi.org/10.1016/j.prevetmed.2005.02.002 [Accessed 23.04.2013].

Reix Nee Broster, C. E., Burn, C. C., Pritchard, J. C., Barr, A. R. & Whay, H. R. (2014). The range and prevalence of clinical signs and conformation associated with lameness in working draught donkeys in Pakistan. *Equine Veterinary Journal* [Online], In press. Available: http://dx.doi.org/10.1111/evj.12231 [Accessed 01. 04. 2014].

Ross, M. W. (2011). Lameness in Horses: Basic Facts Before Starting. *In:* ROSS, M. W. & DYSON, S. (eds.) *Diagnosis and Management of Lameness in the Horse (Second Edition).* 3-8, Available: http://dx.doi.org/10.1016/b978-1-4160-6069-7.00002-x [Accessed 11.02.2014].

Rossel, S., Marshall, F., Peters, J., Pilgram, T., Adams, M. D. & O' Connor, D. (2008). Domestication of the donkey: timing, processes, and indicators. *Proceedings of the National Academy of Sciences of the United States of America* [Online], 105 (10), 3715-20. Available: http://dx.doi.org/10.1073/pnas.0709692105 [Accessed 15.07.2013].

Rousing, T. & Wemelsfelder, F. (2006). Qualitative assessment of social behaviour of dairy cows housed in loose housing systems. *Applied Animal Behaviour Science* [Online], 101 (1-2), 40-53. Available: http://dx.doi.org/10.1016/j.applanim.2005.12.009 [Accessed 14.04.2014].

Roy, R. C., Eager, R., Regan (nee Ashley), F. & Langford, F. (2010). Controlled field trial of a behavioural pain assessment tool in donkeys. *In:* THE BROOKE (ed.) *The 6th International Colloquium on Working Equids: Learning from Others. Proceedings of an International Colloquium held at the India Habitat Centre, New Dehli, India, 29th November to 2nd December 2010.* The Brooke, 30 Farringdon Street, London, EC4A 4HH, UK, 179 - 183, http://www.thebrooke.org/__data/assets/pdf_file/0009/49383/Theme5-Decision-making-in-health-and-disease-91-145.pdf [Accessed 26.07.2013].

Scantlebury, C. E., Aklilu, N., Reed, K., Knottenbelt, D. C., Gebreab, F. & Pinchbeck, G. L. (2013). Ocular disease in working horses in Ethiopia: A cross-sectional study.

Veterinary Record [Online], 172 (4), 99-99. Available: http://dx.doi.org/10.1136/vr.100802 [Accessed 25.11.2013].

Scott, E. M., Nolan, A. M. & Fitzpatrick, J. L. (2001). Conceptual and Methodological Issues Related to Welfare Assessment: A Framework for Measurement. *Acta Agriculturae Scandinavica, Section A - Animal Science* [Online], 51 (sup030), 5-10. Available: http://dx.doi.org/10.1080/090647001316922983 [Accessed 17.06.2013].

Sells, P. D., Pinchbeck, G., Mezzane, H., Ibourki, J. & Crane, M. (2010). Pack wounds of donkeys and mules in the Northern High Atlas and lowlands of Morocco. 42, 219-226. Available: http://dx.doi.org/10.2746/042516409x478532 [Accessed 25.04.2013].

Sim, J. & Wright, C. C. (2005). The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements. *Physical Therapy* [Online], 85 (3), 257 - 268. http://ptjournal.apta.org/content/85/3/257.full.pdf+html [Accessed 20.11.2013].

Sørensen, J. T., Sandøe, P. & Halberg, N. (2001). Animal welfare as one among several values to be considered at farm level: The idea of an ethical account for livestock farming. *Acta Agriculturae Scandinavica - Section A: Animal Science,* SUPPL.**,** 11-16.

Studnitz, M., Jensen, M. B. & Pedersen, L. J. (2007). Why do pigs root and in what will they root? *Applied Animal Behaviour Science* [Online], 107 (3-4), 183-197. Available: http://dx.doi.org/10.1016/j.applanim.2006.11.013 [Accessed 02.04.2014].

Tesfaye, A. & Martin Curran, M. (2005). A longitudinal survey of market donkeys in Ethiopia. *Tropical animal health and production* [Online], 37 Suppl 1, 87-100. Available: http://dx.doi.org/10.1007/s11250-005-9010-5 [Accessed 18.06.2013].

The Donkey Sanctuary (2013). *Our work in Ethiopia* [Online]. Available: http://www.thedonkeysanctuary.org.uk/view/ethiopia [Accessed 19.07.2013 2013].

The Donkey Sanctuary (not stated-a). *Donkey Body Condition Score Chart UK*. Available: http://www.thedonkeysanctuary.org.uk/files/donkeys/BodyConditionScoreChart.pdf [Accessed 21.06.2013].

The Donkey Sanctuary (not stated-b). *A guide to donkey care,* The Donkey Sanctuary, Sidmouth, UK. Available: http://www.thedonkeysanctuary.org.uk/files/donkeys/AGuideDonkeyCare.pdf [Accessed 25.04.2013].

The Donkey Sanctuary (not stated-c). *Our work in Ethiopia* [Online]. Available: http://www.thedonkeysanctuary.org.uk/view/ethiopia [Accessed 28.08. 2013].

Walker, J., Dale, A., Waran, N., Clarke, N., Farnworth, M. & Wemelsfelder, F. (2010). The assessment of emotional expression in dogs using a Free Choice Profiling methodology. *Animal Welfare,* 19**,** 75 - 84.

Welfare Quality® consortium (2009a). *Welfare Quality® Assessment Protocol for Cattle.* Lelystad, The Netherlands.

Welfare Quality® consortium (2009b). *Welfare Quality® Assessment protocol for pigs.* Lelystad, The Netherlands.

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

Wemelsfelder, F., Hunter, T. E. A., Mendl, M. T. & Lawrence, A. B. (2001). Assessing the 'whole animal': a free choice profiling approach. *Animal Behaviour* [Online], 62 (2), 209-220. Available: http://dx.doi.org/10.1006/anbe.2001.1741 [Accessed 05.06.2013].

Wemelsfelder, F., Knierim, U., Schulze Westerath, H., Lentfer, T., Staack, M. & Sandilands, V. (2009a). Qualitative Behaviour Assessment. *In:* FORKMAN, B. & KEELING, L. (eds.) *Assessment of Animal Measurs for Layers and Broilers. Welfare Quality Reports No. 9.* Cardiff: Cardiff University, 113 - 119.

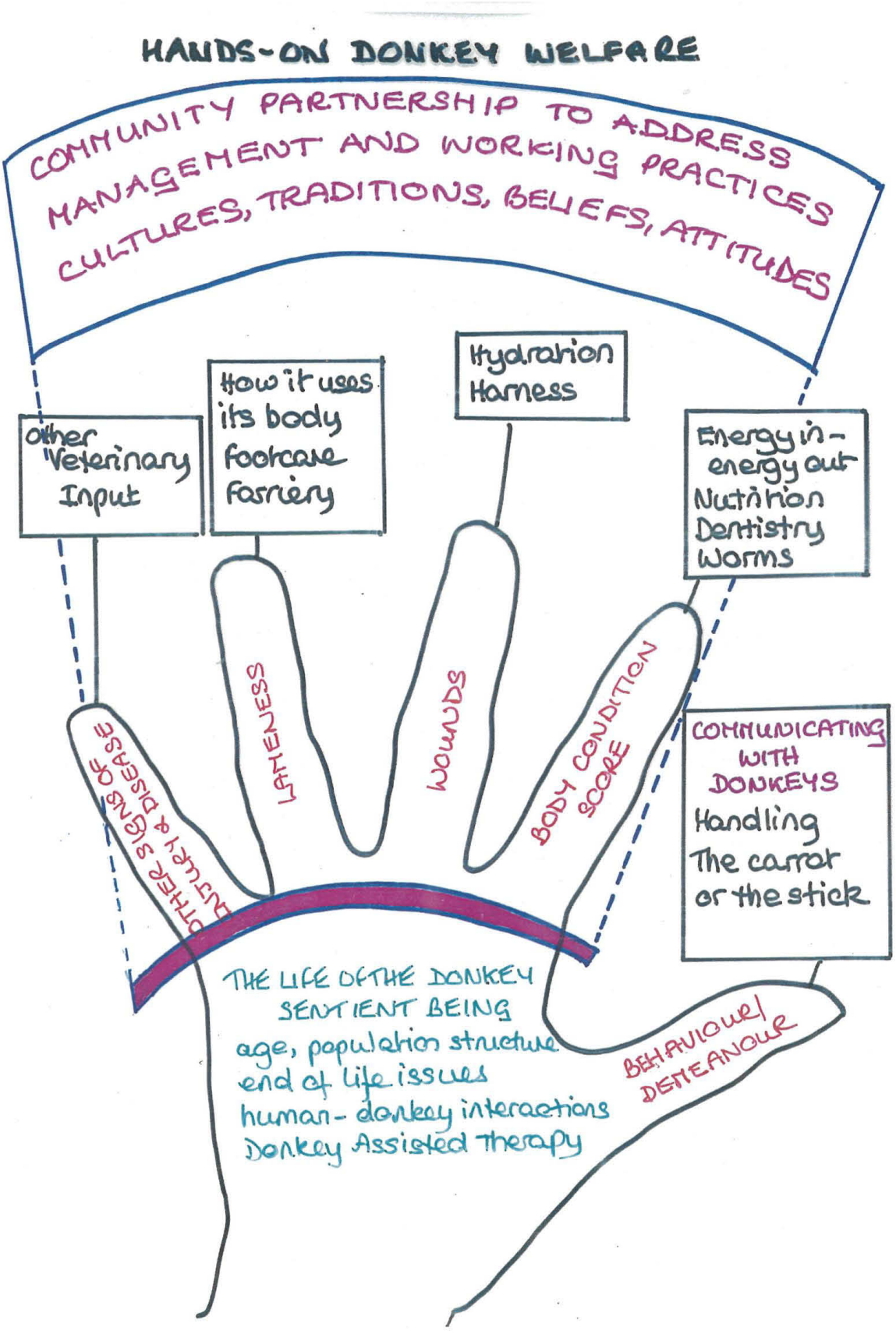
Wemelsfelder, F., Millard, F., De Rosa, G. & Napolitano, F. (2009b). Qualitative Behaviour Assessment. *In:* FORKMAN, B. & KEELING, L. (eds.) *Assessment of Animal Welfare Measures for Dairy Cattle, Beef Bulls and Veal Calves. Welfare Quality Reports No. 11.* Cardiff: Cardiff University, 215 - 224.

Whay, H. R., Main, D. C. J., Green, L. E. & Webster, A. J. F. (2003). Animal-based measures for the assessment of welfare state of dairy cattle, pigs and laying hens: Consensus of expert opinion. *Animal Welfare,* 12**,** 205 - 217.

Wickham, S. L., Collins, T., Barnes, A. L., Miller, D., Beatty, D. T., Stockman, C. A., Blache, D., Wemelsfelder, F. & Fleming, P. A. (2012). Qualitative behavioral assessment of transport-naïve and transport-habituated sheep. *Journal of Animal Science* [Online], 90, 4523 - 4535. Available: http://dx.doi.org/10.2527/jas.2010-3451 [Accessed 16.04.2014].

11 Annex

Original Hands-on Donkey Welfare Protocol



Observer	
Sex	Stallion/Gelding/Mare

General information

Demeanour

General demeanour (0/1/2)	
Observer appr	
Tail tuck (0/1)	
Chin contact (0/1/2/3)	
Ear test (0/1/2)	

Behaviour/Demeanour

BCS

BCS (1/2/3/4/5)	
Teeth (0/1)	Specifiy if 1

Body Condition

Age	

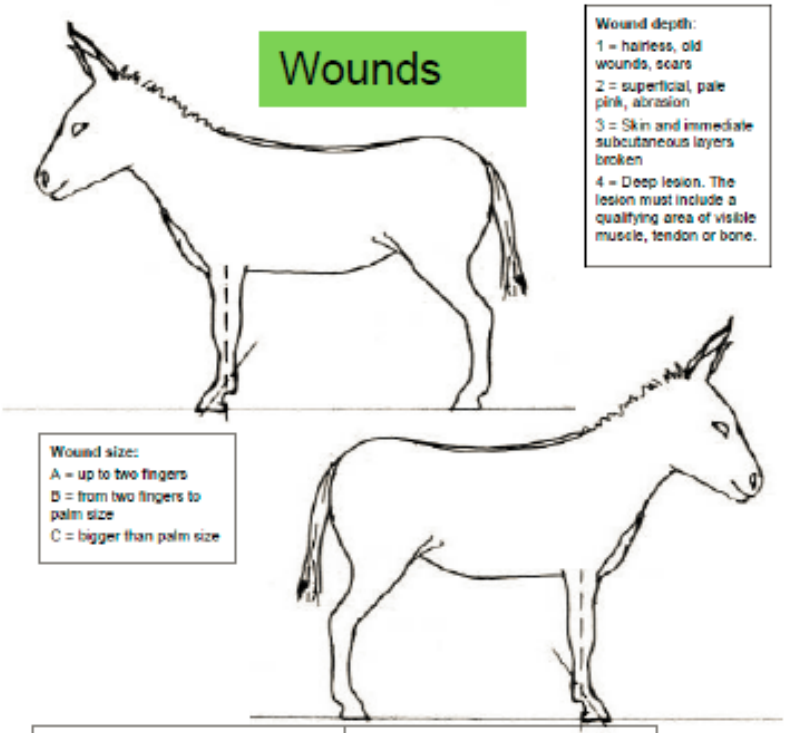
Lameness

Not loaded/loaded	Not tethered/tethered
Weight shifting (0/1)	
Weight bearing (0/1)	
Limb abnormal (0/1)	
Hoof abnormal (0/1)	
Signs of inflammation on the two right limbs (0/1)	
Sole surface of the two right hooves normal (0/1)	
Lame (0/1/2)	

Lameness

Body mapping for any wounds and potentially other problems

If it is a single wound that spans on both sides only note once!



Wounds

Coat condition (0/1/2)	
Evidence of ectoparasites (0	
Heat stress (0/1)	
Discharge from nose (0/1)	
Discharge from eyes (0/1)	
Other signs of diseases	

Other signs of illness and disease

Corrected version of the Hands-on Donkey Welfare Protocol (January 2014)

Observer	
Date _____	Donkey/Mule
Site _____	Market/Village/Other
Work type	Draft / Pack / Ridden / Other
Sex	Stallion/Gelding/Mare

Demeanour

General demeanour (0/1/2)	
Observer approach (0/1/2/3/4)	
Tail tuck (0/1)	
Chin contact (0/1/2/3)	
Ear test (0/1/2)	

BCS

BCS (1/2/3/4/5)	
Teeth (0/1)	Specifiy if 1
Quidding	Missing teeth
Abscess	Hooks

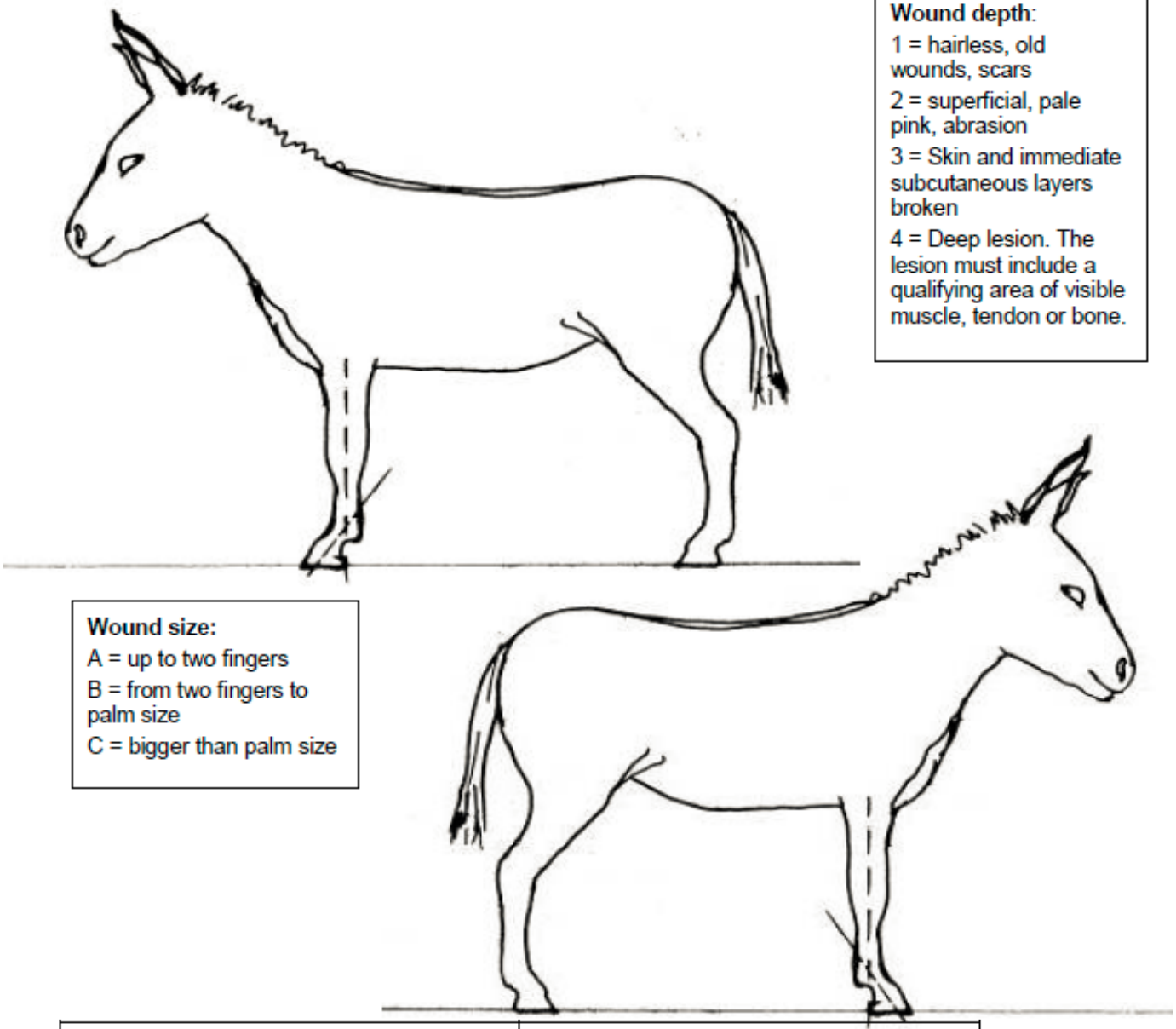
Age	

Lameness

Not loaded/loaded	Not tethered/tethered
Weight shifting (0/1)	
Non Weight bearing (0/1)	
Limb abnormal (0/1)	
Hoof abnormal (0/1)	
Signs of inflammation on the right limb (0/1)	
Abnormal sole surface of the right hoof (0/1)	
Lame (0/1/2)	

Body mapping for any wounds and potentially other problems

If it is a single wound that spans on both sides only note once!



Wound depth:

1 = hairless, old wounds, scars

2 = superficial, pale pink, abrasion

3 = Skin and immediate subcutaneous layers broken

4 = Deep lesion. The lesion must include a qualifying area of visible muscle, tendon or bone.

Wound size:

A = up to two fingers

B = from two fingers to palm size

C = bigger than palm size

Coat condition (0/1/2)	
Evidence of ectoparasites (0/1/2) name	
Heat stress (0/1)	
Discharge from nose (0/1)	
Discharge from eyes (0/1)	
Other signs of diseases	

General aspects

Use one sheet per animal!

To assess the donkey properly (e.g. by touching its legs) you might need a second person to hold the donkey!

1

Behaviour/Demeanour

General demeanour

Before starting to assess the donkey look at its behaviour!

Alert = 0
Active interest in surroundings: looking and listening, ears moving and often forward. Eyes usually wide open. May also show sniffing, vocalising, head movement, tail swishing, moving feet, active interest in other animals and people. Head up unless sniffing or eating.

Apathetic = 1
Passive interest in surroundings: ears may be back but moving in response to stimuli. Eyes may be open or half closed. Responding to self-need, e.g. flapping ears, moving feet or swishing tail to remove flies. Not showing interest in other animals and people unless disturbed by close proximity. Head may be up, level with withers or lowered but raised in response to surrounding activity.

Severely depressed = 2
No interest in surroundings: ears usually back or lowered. Eyes half-closed or fully closed. Minimal or absent response to self-need e.g. flapping ears, moving feet or swishing tail to remove flies. Not roused by passing activity. Not showing interest in other animals and people. Head level with withers or lowered and is not raised in response to surrounding activity.

Source: WEWA

2

Behaviour/Demeanour

Observer approach

Walk at a normal pace towards the animal's head. Approach from an angle of approximately 20 degrees so that it can see you and you are not in its blind spot (directly in front of its face).

0.	**Moves away**. Animal moves or attempts to move sideways or backwards away from you.
1.	**Turns head away.** The animal only turns its head away and does not try to move away as well.
2.	**Does not move.**
3.	**Turns head towards observer.** This should look like a sign of friendly interest.
4.	**Aggressive:** The animal does one or more of the following. a) turns toward you with the intention of biting, rearing or striking out with a foreleg, b) moves or turns towards you with ears held back or flattened against its head, c) spins around or kicks out towards you with its hind legs.

Source: adapted from WEWA

3

Behaviour/Demeanour

Chin contact test

Gently put one hand under the chin and make contact until your hand is lightly taking some weight, but not lifting the head. If the animal moves its head to avoid the approaching hand, do not pursue it. Assess the animal's response at the moment of attempting to make contact. Only make one attempt to contact the chin.

0	**Accepts chin contact**. The animal quietly allows its chin to be touched.
1	**Avoids chin contact or withdraws head when chin contacted.** The animal may throw its head up or toss it sideways as hand is approaching or when touched. Gentle avoidance of chin contact
2	**Exaggerated avoidance of chin contact**
3	**Total refusal of chin contact.** Unable to score chin contact, impossible to assess.

Ear contact test

After having put the hand under the chin, touch the hand to the ear.

0	**tolerates ear touch.** The donkey allows the ear to be touched.
1	**avoids ear touch.** The donkey moves or shies away from the tactile experience.
2	**Total refusal of ear contact.** Unable to score ear contact, impossible to assess.

4

Source: adapted from WEWA and McLean 2012

Behaviour/Demeanour

Tail tuck

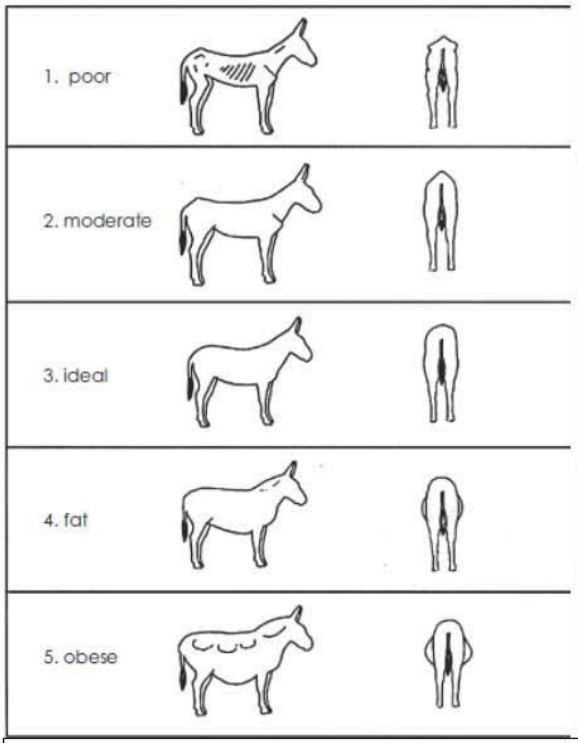
Note if the donkey tuck its tail during the observer approach or when you walk towards its hind limbs.
(Do not count if the donkey tail its tuck when you look for tail base wounds and lift up its tail!)

0	**Does not tuck its tail.**
1	**Tuck its tail.**

5

Source: adapted from WEWA and McLean 2012

BCS



C/S	NECK AND SHOULDERS	WITHERS	RIBS AND BELLY	BACK AND LOINS	HINDQUARTERS
1. POOR	Neck thin, all bones easily felt. Neck meets shoulder abruptly, shoulder bones felt easily, angular.	Dorsal spine of withers prominent and easily felt.	Ribs can be seen from a distance and felt with ease. Belly tucked up.	Backbone prominent, can feel dorsal and transverse processes easily.	Hip bones visible and felt easily (hock and pin bones). Little muscle cover. May be cavity under tail.
2. MODERATE	Some muscle development overlying bones. Slight step where neck meets shoulders.	Some cover over dorsal withers, spinous processes felt but not prominent.	Ribs not visible but can be felt with ease.	Dorsal and transverse processes felt with light pressure. Poor muscle development either side midline.	Poor muscle cover on hindquarters, hipbones felt with ease.
3. IDEAL	Good muscle development, bones felt under light cover of muscle/fat. Neck flows smoothly into shoulder, which is rounded.	Good cover of muscle/fat over dorsal spinous processes withers flow smoothly into back.	Ribs just covered by light layer of fat/muscle, ribs can be felt with light pressure. Belly firm with good muscle tone and flattish outline.	Cannot feel individual spinous or transverse processes. Muscle development either side of midline is good.	Good muscle cover in hindquarters, hipbones rounded in appearance, can be felt with light pressure.
4. FAT	Neck thick, crest hard, shoulder covered in even fat layer.	Withers broad, bones felt with firm pressure.	Ribs dorsally only felt with firm pressure, ventral ribs may be felt more easily. Belly overdeveloped.	Can only feel dorsal and transverse processes with firm pressure. Slight crease along midline.	Hindquarters rounded, bones felt only with firm pressure. Fat deposits evenly placed.
5. OBESE	Neck thick, crest bulging with fat and may fall to one side. Shoulder rounded and bulging with fat.	Withers broad, unable to feel bones.	Large, often uneven fat deposits covering dorsal and possibly ventral aspect of ribs. Ribs not palpable. Belly pendulous in depth and width.	Back broad, unable to feel spinous or transverse processes. Deep crease along midline bulging fat either side.	Cannot feel hipbones, fat may overhang either side of tail head, fat often uneven and bulging.

Aged donkeys can be hard to condition score due to lack of muscle bulk and tone giving thin appearance dorsally with dropped belly ventrally, while overall condition may be reasonable.

6

BCS 1

BCS 2

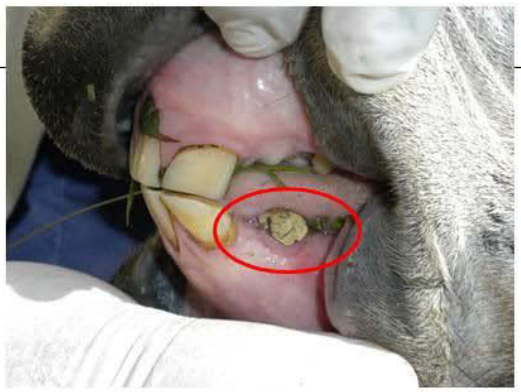
7

BCS

Teeth

Molars can be basically assessed by running the hands up and down. Then scratch the teeth towards each other and listen for abnormal sounds.

0	**Normal.**
1	**Abnormal.** Specify the abnormality: Missing teeth Quidding Abscess Hooks



8

BCS

Age

While looking at the teeth estimate the animal's age.

0	0 to 5 years
1	6 to 10 years
2	11 to 15 years
3	> 16 years

9

Age

0 – 5 years

2,5 years = central incisor teeth present
3,5 years = lateral incisor teeth present
4,5 years = corner incisor teeth erupting

6 – 10 years

Oval dental cups
Angle of upper and lower teeth around 90 %

11 – 15 years

Round dental cups
Angle becomes more acute

> 16 years

Acute angle of upper and lower teeth, no dental cups

10

Lameness

First state if the donkey is loaded and/or tethered!

Weight shifting

If the donkey is standing look if it is shifting its weight from one leg to the other not wanting to bear weight on one or more legs too long (Note: this is usually of importance when the donkey is loaded.).

0	**Absent**
1	**Present**

11

Lameness

Non weight bearing

The donkey is not bearing its weight evenly on all four legs.

0	**Absent**
1	**Present**

Lameness

Whole limb abnormal

Look for fractures, dislocations, Carpus varus/valgus, cow hocked hind legs, obvious swellings (not including the signs of inflammation), toe in/out etc.

0	Absent
1	Present

13

Lameness

Hoof abnormal

hoof too long or too short, hoof cracked, problems at coronary band

0	Absent
1	Present

Hoof cracked and problems at coronary band

Hoof too long

14

Lameness

Signs of inflammation on limb and/or hoof

Go along the two limbs on the right side (from elbow/knee to hoof) and look for signs of swelling, heat, pain, redness and loss of function.

0	Absent
1	Present

Abnormal sole surface

When coming to the hooves of the two legs on the right side lift them up and look if the sole is uneven and abnormal.

0	Absent
1	Present

15

Lameness

Lame

If the donkey shows one of the indicators stated above score it as *potentially lame*. If you can see the donkey walking lame then score *obviously lame*.

0	Not lame
1	Potentially lame
2	Obviously lame

16

Wounds

If there are not clearly distinguishable old wounds present (white hair) on the legs, then score 1C.

3 A

17

Other signs of illness and disease

Coat condition

Also have a look on the neck as the coat on the back could be assessed incorrectly due to sweat or dirt.

0	**Excellent.** Coat shiny, glossy, flat, clean and unbroken
1	**Average**. Coat not shiny. A few minor abrasions acceptable.
2	**Poor.** Coat dull, staring, matted, dry and uneven

1

0

18

Other signs of illness and disease

Signs of ectoparasites

Look under the tail (ticks), neck and chest (Gasterophilus eggs, lice) for signs of ectoparasites, flies on legs.

0	No signs on body
1	Signs for little to moderate infestation
2	Signs for heavy infestation

19

Other signs of illness and disease

Heat stress

If **3 out the following 5 signs are shown then heat stress is present: flared nostrils, apathy, increased, respiratory depth, increased respiratory rate, head nodding**

0	absent
1	present

Discharge from nose and eyes

0	absent
1	present

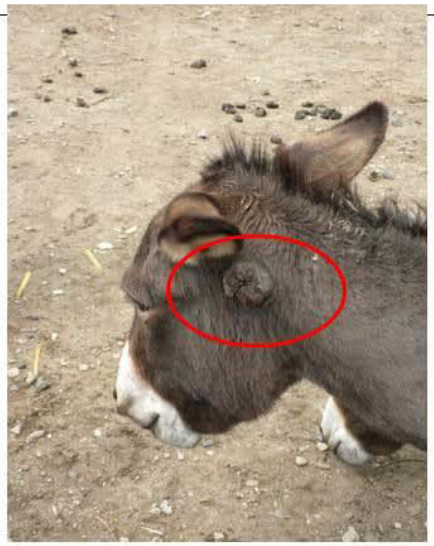
20

Other signs of illness and disease

University of Natural Resources and Life Sciences, Vienna

Other signs of diseases (swelling of other body parts than legs, oedema, sarcoid etc.)

Please state!



Sarcoid

21