Welfare® Quality
Science and society improving animal welfare

Welfare Quality Reports No. 11

Assessment of Animal Welfare Measures for Dairy Cattle, Beef Bulls and Veal Calves

edited by
B. Forkman
L. Keeling

Welfare® Quality

Science and society improving animal welfare

Assessment of Animal Welfare Measures for Dairy Cattle, Beef Bulls and Veal Calves

edited by

B. Forkman and L. Keeling

SIXTH FRAMEWORK PROGRAMME

CARDIFF UNIVERSITY PRIFYSGOL CAERDYDD

Welfare Quality Reports No. 11

Welfare Quality Reports
Edited by Mara Miele and Joek Roex

School of City and Regional Planning
Cardiff University
contact: LeoR@cardiff.ac.uk

September 2009

The present study is part of the Welfare Quality® research project, which has been co-financed by the European Commission, within the 6th Framework Programme, contract No. FOOD-CT-2004-506508. The text represents the authors' views and does not necessarily represent a position of the Commission who will not be liable for the use made of such information.

Typeset by Joek Roex, Cardiff.
Printed in Sweden by SLU Service/Reproenheten, Uppsala.

ISBN 1-902647-81-5
ISSN 1749-5164

CONTENTS

PREFACE

The current book contains a number of chapters based on the work done by scientists working on the work package 2.2 of the EU-project Welfare Quality® 'Integration of animal welfare in the food quality chain: from public concern to improved welfare and transparent quality'. One aim of Welfare Quality® has been to develop a scheme for assessing animal welfare.

The aim of the current work package was to test and validate possible measures of animal welfare in such a scheme for poultry, pigs and cattle. This book deals with the measures on cattle. The measures included in this book have been assessed for dairy cows, fattening bulls and veal calves.

Welfare Quality® has identified four welfare principles; good feeding, good housing, good health and appropriate behaviour. These four welfare principles has been further divided into twelve welfare criteria (see Table). The focus has been on animal based measures, while resource based measures and management based measures have been used to supplement these. The areas covered in the lives of the animals are on farm and at slaughter. The transport is only covered in the loading/unloading phase, because the effect of the transport can often be seen in the state of the animals at the slaughter plant.

The measures evaluated have been of three categories. The first category are measures that have not previously been validated, the second category is measures that have been validated, or have high face validity (there is e.g. no need to do a study to validate that wounds cause bad welfare) but for which we do not know the repeatability of the measure. The last category finally consists of measures for which there are already validated protocols (e.g. for body condition score), in these cases the project groups evaluated the alternatives and selected the most appropriate measure to be used.

Welfare principles	Welfare criteria
Good feeding	1 Absence of prolonged hunger 2 Absence of prolonged thirst
Good housing	3 Comfort around resting 4 Thermal comfort 5 Ease of movement
Good health	6 Absence of injuries 7 Absence of disease 8 Absence of pain induced by management procedures
Appropriate behaviour	9 Expression of social behaviours 10 Expression of other behaviours 11 Good human-animal relationship 12 Positive emotional state

This book is based on the work of different research groups, each chapter therefore deals with related aspects of animal welfare, e.g. 'health scores', which covers 15 potential measures of the health of the animal. This means that although there are only 23 groups of measures/chapters, the number of measures evaluated is much higher since each group contains information on a number of different measurements.

The current book consists of all of the measures that have been evaluated for inclusion in the welfare assessment scheme of Welfare Quality (whether they were later included or not). In addition we have included two appendices on possible management and resource based measures. However, these have not been tested in the current project. Finally, there is a third appendix with a brief note on statistics and sampling.

Described below is how each of chapters relate to the twelve animal welfare criteria.

1. Absence of prolonged hunger. The proposed measure for prolonged hunger in cattle was condition scoring (Chapter 1).
2. Absence of prolonged thirst. There is currently no good animal based measure for this criterion. The resource and management based measure suggested is the number of water cups and that they are functioning (Appendix 1 and 2).
3. Comfort around resting. The hypothesis that the behaviour around resting can be used as a measure of the comfort of dairy cattle and bulls was investigated (Chapter 2). The cleanliness of the animals is also an indicator of comfort around resting (Chapter 3). This could also possibly be measured at the slaughter plant (Chapter 4)
4. Thermal comfort. No animal-based measure of thermal comfort on farm was investigated.
5. Ease of movement. The main restriction the animals experience is the either the use of tie stalls, or because the animals are densely stocked, this concern has therefore been addressed in the resource appendix.
6. Absence of injuries. There are two main types of injuries investigated for this welfare criteria. The first one is lameness (Chapter 5). The second one is integument alterations (Chapter 6). Included in this type of injury are swellings and skin condition, including udders (in dairy) and prepuce (in bulls). Bruises may also be discovered on the carcasses at slaughter (Chapter 7). The category of injurious behaviours (slipping, stepping on other individuals etc.) has also been investigated for possible inclusion in the final scheme (on farm: Chapter 8, at slaughter: Chapter 9).
7. Absence of diseases. Respiratory, enteric, and reproductive problems can be assessed on farm in a valid and feasible way (Chapter 10). Dead on arrival at the slaughter plant is another possible candidate for monitoring the health of the animals (Chapter 4)
8. Absence of pain induced by management procedures. The pain caused by mutilation (e.g. dehorning) is most easily recorded as a management measure (see the management appendix). At slaughter the stunning effectiveness is regarded as an important measure within this area of concern (Chapter 11).
9. Expression of social behaviour. Social behaviour can both be agonistic and cohesive. The use of direct registrations of agonistic behaviour (as opposed to wounds) was

investigated (Chapter 12). In cattle the possibility of measuring social licking as a cohesive behaviour was investigated (Chapter 13).

10. Expression of other behaviours. For the welfare criteria ‘expression of other behaviours’ several types of behaviours were investigated. There are behaviours that indicate poor welfare, e.g. abnormal behaviours, including stereotypes (Chapters 8 and 14).
11. Good human–animal relations. The avoidance reaction of the cattle can be used to measure the human animal reaction (Chapters 15, 16, and 17).
12. Positive emotional state: Possible indicators of positive emotional state is exploration and play behaviour (Chapters 18 and 19). Fearfulness is the opposite of a positive emotional state. The measurements of fear on farm (Chapter 21 and 22) and at slaughter (Chapter 9) were investigated. Finally, it is possible to gather information concerning the overall emotional state of the animal as it is perceived by means of Qualitative Behaviour Assessment (Chapter 23).

B. Forkman
September 2009

LIST OF CONTRIBUTORS

B. Algers
Swedish University of Agricultural Sciences, Skara, Sweden

E.A.M. Bokkers
Wageningen University and Research Centre, Lelystad, Netherlands

X. Boivin
INRA, Clermont-Ferrand, France

N. Brörkens
University of Kassel, Germany

E. Canali
University of Milan, Italy

Antoni Dalmau
IRTA, Spain

G. De Rosa
University of Naples Federico II, Italy

B. Forkman
University of Copenhagen, Denmark

L.F.M. Heutinck
Wageningen University and Research Centre, Lelystad, Netherlands

L.J. Keeling
Swedish University of Agricultural Sciences, Uppsala, Sweden

U. Knierim
University of Kassel, Germany

S. Laister
University of Natural Resources and Applied Life Sciences, Vienna, Austria

K.A. Leach
University of Bristol, United Kingdom

B.J. Lensink
Wageningen University and Research Centre, Lelystad, Netherlands

H. Leruste
Institute Supérieur d'Agriculture, Lille, France

S. Lolli
University of Milan, Italy

N. MacKintosh
University of Natural Resources and Applied Life Sciences, Vienna, Austria

F. Millard
Scottish Agricultural College, Edinburgh, United Kingdom

M. Minero
University of Milan, Italy

F. Napolitano
University of Basilicata, Potenza, Italy

G. Plesch
University of Kassel, Germany

R. Quast
University of Kassel, Germany

A.-M. Regner
University of Natural Resources and Applied Life Sciences, Vienna, Austria

H. Schulze-Westerath
Institute of Animal Sciences, Zürich, Switzerland

C. Schmied
University of Veterinary Medicin, Vienna, Austria

J.T.N. van der Werf
Wageningen University and Research Centre, Lelystad, Netherlands

C.G. van Reenen
Wageningen University and Research Centre, Lelystad, Netherlands

Antonio Velarde
IRTA, Spain

S. Waiblinger
University of Veterinary Medicin, Vienna Austria

F. Wemelsfelder
Scottish Agricultural College, Edinburgh, United Kingdom

R. Westin
Swedish University of Agricultural Sciences, Skara, Sweden

H.R. Whay
University of Bristol, United Kingdom

C. Winckler
University of Natural Resources and Applied Life Sciences, Vienna, Austria

I. Windschnurer
University of Veterinary Medicin, Vienna, Austria

M. Wolthuis-Fillerup
Wageningen University and Research Centre, Lelystad, Netherlands

K. Zegner
University of Natural Resources and Applied Life Sciences, Vienna, Austria

D. Zucca
University of Milan, Italy

1

CONDITION SCORING FOR DAIRY AND BEEF CATTLE AND VEAL CALVES

K.A Leach, U. Knierim and H.R. Whay

1.1 SUMMARY

Body condition reflects the body fat content and thus the nutritional status of an animal, as indicated by its body reserves. Body condition has important implications for health. There can be problems of dystocia and fatty liver disease, and delayed return to oestrus, with over-fat dry cows. In early lactation, high producing dairy cows can lose a great deal of condition due to their negative energy balance, and this can have a detrimental effect on health and fertility. Poor body condition may reflect previous prolonged hunger or ill health.

Published body condition scoring systems were reviewed, many of which require palpation, and are very detailed. Since palpation will not be feasible for beef animals in welfare assessments, we propose a simplified system, based on a published scale, but using visual assessment of the loin and tailhead areas. This will identify animals which are too thin, and also too fat in the case of dairy cattle. A representative random sample of the dairy herd, including dry cows, should be scored. Beef cattle will often be more difficult to view than dairy cows. For these we suggest that, if random sampling (to give a reliable prevalence figure for thin animals), is not possible, the assessor should view all groups as thoroughly as possible and any animals classified as 'too thin' should be reported. There are risks of either double counting or false negatives with this method. More than a certain number of animals which are 'too thin' should trigger a second, more detailed investigation. No condition scoring methods have been validated for veal calves to date. It would be possible to use the same system for assessing calves, but the validity and implications of condition score in young calves are not known.

1.2 INTRODUCTION

Body condition reflects the body fat content (Wagner et al., 1988) and thus the nutritional status of an animal, as indicated by its body reserves. It reflects the nutritional history of the animal, rather than the current nutrition. Body condition has important implications for health. For example, there can be problems of dystocia and fatty liver disease and delayed return to oestrus with over-fat dry cows (Reid et al., 1986), while in early lactation, high producing dairy cows can lose a great deal of condition due to their negative energy balance, and this can have a detrimental effect on health and fertility (Butler, 2003). Low body condition per se is an indicator that the energy needs of the animal are not being met by the dietary energy supply. Body condition monitoring is quite widely used as a tool in management of both beef and dairy cattle. The purpose of including body condition scoring within a welfare assessment would be to identify the proportion of animals that are either too thin or too fat, indicating increased risk of disease, and/or inappropriate former and possibly current nutrition. Once these animals have been identified, their management can be improved.

A wide range of body condition scoring systems have been developed and used for research purposes and practical monitoring on commercial farms. These were reviewed and discussed in order to decide upon a suitable method.

1.3 METHODS

The history of the development of body condition scoring was reviewed. The various scoring systems in use were considered in terms of their feasibility, validity, simplicity and reliability. The suitability of existing methods, and the possibility of creating a new method were considered, in the context of farm assessment of welfare. The practicalities of applying the chosen system to dairy, beef and veal calves were addressed. The proposed system was tested for reliability between observers by three observers scoring the same set of photographs. Two were experienced in condition scoring and one was inexperienced, and was given the training materials to study before being asked to score the pictures.

1.4 RESULTS AND DISCUSSION

Various condition scoring systems have been developed in different countries. The main distinctions between systems are:

- Whether they are merely visual or require palpation, and
- Whether the animal is assessed as a whole, or separate scores are given for different anatomical regions, which are then summarised or adjusted to give a whole animal score.

The most commonly used and familiar systems for condition scoring in the UK use a six point scale from 0 to 5, and are based on the work of Lowman et al (1974) (who published the first system for beef cattle), and Mulvany et al. (1977) who applied the system to dairy cows. These systems were produced as tools for monitoring the condition of cows at different stages of the production cycle, and aiding management. Slightly different descriptions were created for the two types of cattle, related to their tendency to deposit fat at different anatomical locations at the higher levels of fat deposition. These two systems were designed to include palpation of the tailhead and loin areas, however, notes on Mulvany's system suggest that palpation is only necessary for the refinement of allocating half scores. In fact, scoring by visual assessment alone is commonly practiced.

Condition score assessment in North America has developed along slightly different lines, with greater divergence between approaches for dairy cattle and beef cattle. The earliest publications were of descriptive methods for beef cattle, generally a nine point system requiring both visual assessment and palpation (e.g. Richards et al., 1986; Wagner et al., 1988). Edmondson et al. (1989) developed and validated a system for dairy cows, based on visual assessment. This method requires consideration of separate regions of the body, but a whole animal score is finally allocated. The structured technique of assessing different areas of the body can be useful in training, but the method is quite detailed since quarter scores are included. The combination of assessing a large number of body areas, with a large number of possible scores, makes the system rather complex. Fergusson et al. (1994) carried out an experiment to define the principal components of a system for dairy cows which was similar to that of Edmondson et al. (1989), in that seven body regions were separately assessed, and given a score between 1 and 5, with quarter scores, based on simple text descriptions of the appearance of the underlying bone structures. Although the descriptions were visual, tactile assessment was also permitted. Principal components analysis was then used to determine the particular criteria for individual cut-off points. For example, this analysis showed that the appearance of the thurl region (U or V shape) was the critical component separating scores above or below 3, and the appearance of hook and pin bones determined whether cows were at or above 2.75. These results allowed the development of a scheme or key system, versions of which are used by many of the American state agricultural advisory services, e.g. <http://cahpwww.vet.upenn.edu/dairy/bcs_cht.htm>.

Both the UK-style and American-style systems have been widely used in research, for investigating relationships between body condition, or its change, and many aspects of health, fertility and production. However, as already mentioned, they are also used at a practical level, providing targets for optimum condition at certain stages of the production cycle, particularly for dairy and suckler cows. For fattening beef cattle they are used more in relation to carcass composition.

For the purposes of this welfare assessment, the ability to detect animals which are at an inappropriate condition is the main objective. This means that a simple classification of 'too thin', 'acceptable' and 'too fat' will suffice. Using the same description for the condition score classes for both beef and dairy animals would simplify both the training and assessment procedures. However, slightly different thresholds were eventually decided upon for dairy breeds, compared with dual purpose breeds and beef breeds, to recognise that condition loss is less likely, and therefore more of a concern, in dual purpose or beef breeds. A method requiring palpation was not considered feasible, particularly for beef animals, therefore a system which could be operated using only visual assessment was needed.

In terms of welfare, it is generally accepted among users of the UK condition scoring system that there is cause for concern when any animals fall below condition score 2 and when pregnant dry cows reach condition score 4 or above (Ward, 2003). The American systems generally suggest cause for concern at scores which equate closely to these cut-off points. A simple description of the features which characterise animals above and below these cut-off points, based on the UK descriptions, was therefore derived (Tables 1.1 and 1.2). This largely agreed with descriptors of equivalent differential points from other systems. The indicators for all four body regions should be present to classify an animal as 'too thin' or 'too fat'.

Although there can be discussion about the capacity and necessity of dairy cows to lose body condition in early lactation, the condition identified by this system as 'too thin' (i.e. condition score less than 2 according to Lowman et al. (1973) or Edmondson (1989) should be universally unacceptable for any animal at any stage of lactation.

The proportion of animals in a group classified as 'too fat' or 'too thin' should therefore be determined. Animals for sampling within a dairy herd could be selected by the same methods as for clinical scoring and condition scoring. The same sample of animals could

TABLE 1.1 Classification of dairy breeds as too thin and too fat.

Region	Too thin (Lowman/Mulvany score <2, Edmondson et al. score <2.5)	Too fat (Lowman/Mulvany/ Edmondson et al. score 4 or more)
Tailhead	Cavity around tailhead	Tailhead cavity full and folds of fatty tissue present
Loin	Deep depression between backbone and hip bones (tuber coxae)	Convex between backbone and hip bones (tuber coxae)
Vertebrae	Ends of transverse processes sharp	Transverse processes not discernible
General	Tailhead, hip bones (tuber coxae), spine and ribs prominent	Outlines of fat patches visible under skin

TABLE 1.2 Classification of of beef and dual purpose breeds as too thin and too fat.

Region	Too thin (Lowman/Mulvany score <3, Edmondson et al. score <3.25)	Too fat (cows in dairy systems only) (Lowman/Mulvany/ Edmondson et al. score 4 or more)
Tailhead	Cavity around tailhead	Tailhead cavity full and folds of fatty tissue present
Loin	Visible depression between backbone and hip bones (tuber coxae)	Convex between backbone and hip bones (tuber coxae)
Vertebrae	Ends of transverse processes distinguishable	Transverse processes not discernible
General	Tailhead, hip bones (tuber coxae) visible	Outlines of fat patches visible under skin

be used. Power calculations will indicate the number of cows that need to be scored to estimate the prevalence of a condition (e.g. 'too fat' or 'too thin') within a group of a certain size (finite population), with a set confidence interval (see Appendix 3). A minimum number of cows to score should be set, to ensure that small herds are properly assessed – we suggest a minimum of 30 cows. If there are different management groups they should be sampled proportionally. Dry cows should definitely be sampled and assessed.

The random sample of cows may be selected by marking every rth cow encountered while walking through the housing, in the parlour, or while they are restrained at the feed trough (r being calculated from the herd size and power calculation determining the number of animals required: ie r = N/n where N is herd size and n is sample size). Care should be taken not to exclude animals which are lying down. The same sample of animals may be used for other animal-based assessments, such as clinical indicators and cleanliness. Alternatively, animals for the sample may be identified from a herd list, and observed within or as they leave the milking parlour, or every nth cow entering the parlour could be assessed. However, this will commit the assessor to being present for the whole of the milking time. This may be efficient for smaller herds.

If all the animals in a group have been restrained for examination, animals can be released individually once they have been scored. If they are already loose, their numbers must be recorded as they are scored, or their identity determined, for example by a second mark, to prevent scoring the same animal twice. Where there are a number of different groups of cows, care must be taken that all are proportionally represented in the herd sample.

Beef cattle are likely to be kept in conditions where they cannot be closely controlled, and in this situation, only a group level assessment will be feasible. In this case, for each group, detection of any animals which are 'too thin' should be recorded. However, there are risks of either double counting or false negatives with this method.

After development of the current system, two experienced observers scored the same photographs using the new system, with percentage agreement of 90% for beef cattle and 79% for dairy cows. Including the observations of an inexperienced observer, trained only using photographs, reduced the percentage agreement to 67% and 65% for the two groups of animals respectively. This suggests that training on live animals should definitely be included.

Body condition assessment for veal calves has not been validated, therefore it is not possible to make definitions and recommendations for target condition scores.

1.5 CONCLUSIONS

Body condition for welfare purposes can be assessed with a simple three category system of ‘too thin’, ‘acceptable’ and ‘too fat’. These categories are differently defined for dairy and dual purpose or beef breeds. Dairy cows can be scored individually, with a random representative sample being taken. Beef cattle are likely to be less easy to score individually. Identification of any animals which are too thin should be attempted, although there is a risk of either double counting or underestimation. Body condition assessment for veal calves has not been validated, therefore it is not possible to make definitions and recommendations for target condition scores

2

RELIABILITY TESTING CONCERNING BEHAVIOUR AROUND RESTING IN CATTLE IN DAIRY COWS AND BEEF BULLS

N. Brörkens, G. Plesch, S. Laister, D. Zucca, C. Winckler, M. Minero and U. Knierim

2.1 SUMMARY

Lying is a high priority behaviour and disturbances around resting are highly welfare relevant as they may be associated with insufficient recuperation and frustration, increased risks for lameness and alterations or injuries regarding hair, skin and joints. A wide variety of different measures of behaviour around lying has been used in the literature. Some of the measures are not applicable within a short-term on-farm welfare assessment. For other measures information about their suitability as on-farm welfare measures is limited.

It was the aim of this study to investigate in dairy cows and beef bulls these behavioural measures around resting with regard to their feasibility, inter-observer reliability and short- to long-term intra-farm variability (consistency). For this purpose, observations on 35 dairy (16 cubicle houses, 7 deep litter systems, 12 tie stalls) and 19 beef farms (9 deep litter, 10 fully slatted floors) were carried out on three days for about 5 h each. Farm visits took place at approximately 60 and 180 days (tie stalls: 120 days) after the first visit. Additionally, inter-observer reliability was tested in direct observations on 5 farms, from video clips with 65 lying down occurrences and from 57 pictures with in total 67 lying animals.

Many categories of behaviour around resting occurred so infrequently that their reliable recording is questionable with regard to a short-term observation of about 2 hours. Inter-observer reliability was generally good (Spearman r_s or Kendall's W = 0.75-1.00) for those measures that could be recorded more than once per observation hour, except for the percentage of collisions during rising or lying. As this measure showed a good consistency of results over time (Kendall's W=0.88), it is suggested to improve training and investigate inter-observer reliability again in the next stage of the project. In general, only a small

number of measures showed an acceptable consistency of results over time, especially when observation times of one or two hours were simulated.

We finally recommend to include into an on-farm welfare monitoring system for dairy cows the measures ‘duration of lying down’, taking the total duration of at least 6 voluntary occurrences, and the ‘percentage of collisions during lying down’ during these occurrences. Additionally, during the first two hours after the morning feeding, the ‘percentage of cows lying partly or completely outside the lying area’ should be recorded by instantaneous scan sampling over the whole pen every 10 to 20 minutes. During the two hours of observations, further behavioural measures can be recorded. In beef bulls, a new measure ‘ratio of lying bulls ruminating to all bulls ruminating’ is proposed that should be tested and evaluated in further stages of the project. Furthermore, ‘duration of lying down’ can reliably be recorded, if a minimum of 8 voluntary occurrences in bulls heavier than 350 kg is achieved. This might take on average 5 hours. However, during this time further data can be collected.

2.2 INTRODUCTION

Lying is a high priority behaviour (Munksgaard et al., 2005) and cattle rest mainly while lying. Disturbances of the behaviour around resting may be associated with insufficient recuperation and frustration (Munksgaard and Simonsen, 1996), increased risks for lameness (Singh et al., 1994) and alterations or injuries regarding hair, skin and joints (Wechsler et al., 2000). A number of different measures around resting have been used in past studies with regard to beef bulls and dairy cows. While some of these measurements such as total lying time or number lying periods cannot be recorded in short-term observations, others may be suitable for the on-farm welfare assessment. Among them are:

- time needed to lie down and to get up (durations of rising and lying down)’
- percentage of animals colliding with housing equipment during rising or lying down;
- percentage of animals with altered or abnormal rising or lying down behaviour;
- percentage of animals slipping during lying down or rising;
- percentage of interrupted lying down or rising movements (unsuccessful lying down or rising attempts);
- percentage of animals in the different lying positions;
- percentage of cattle ruminating during lying;
- synchrony of lying;
- percentage of animals lying partly or completely outside lying area (only in dairy cows);
- percentage of animals standing in lying area (only in dairy cows in loose housing).

It was the goal of this study to apply these measures on dairy and beef bull farms in order to investigate their feasibility, inter-observer reliability and the consistency of results over time.

2.3 METHODS

2.3.1 Inter-observer Reliability Testing (IORT)

During several meetings between partners, detailed definitions of the measures were developed and intensive training including reliability testing undertaken. Formal inter-observer reliability testing took place during the following sessions:

1. July 2005 (Germany), before start of regular observations, 2 observers (A, B): on-farm observations of 21 rising occurrences and video observations of 34 lying down occurences
2. September 2005 (Austria), 55 days after first observations, 2 observers (A, B): on-farm recordings of 20 scans with in total 51 lying animals, 6 rising and 13 lying down occurrences on 2 dairy farms.
3. June 2006 (Italy), 310 days after first observations, 3 observers (A, B, C): on-farm recordings of 30 scans with in total 35 lying animals, 8 rising and 21 lying down occurrences on 2 dairy farms.
4. July 2006 (separately at different places), around 400 days after first observations, 2 observers (A, B): video observations of 65 lying down movements and scoring of 57 pictures with in total 76 lying animals

In test session 3 a third observer from Italy participated in the reliability testing. For test session 4, pictures of animals in different lying positions and videos of animals lying down had been taken during the earlier farm visits from about the same position as for the on-farm observations. Pictures and videos showed one or several animals kept in different housing conditions.

No distinction was made between dairy cows and beef bulls as the behaviour patterns are the same in both categories of cattle.

Depending on the number of observers, Spearman (r_s) or Kendall's W correlation analyses were performed. For nominal data , the prevalence adjusted bias adjusted kappa coefficient PABAK was calculated.

2.3.2 On-farm Recordings

Observation of resting behaviour on 19 beef bull farms (9 deep litter including sloped straw flow systems and 10 fully slatted systems, 30–220 animals per farm, 5-27 bulls per pen) and 35 dairy farms (16 cubicle houses, 7 deep litter including sloped straw flow systems, 12 tie stall systems; herd size 12–100 cows) were carried out in Germany and Austria (Table 2.1) on three days to test short-, medium- and long-term variability of the measures within farms (see Figure 2.1) using Kendall's W correlation analysis. All observations took place from August 2005 until April 2006.

Farm visits took place 60 and 180 days (± 10 days; one farm on day 227 instead of 180) after the first visit (beef cattle farms, dairy loose housing systems). In dairy farms with tie stalls, the third visit took place 120 days after the first visit because of organisational reasons.

In dairy cows, separate groups of dry or periparturient cows or cows in hospital pens were not observed. In herds larger than 25 cows, the observations were carried out in segments of the barn which were expected to contain on average not more than 25 cows per segment. This was mainly done because of the concurrent observation of other behaviours (e.g. social behaviour, play) by behaviour sampling which sets an upper limit to the number of animals that can be observed at one time. However, also scan sampling was difficult to perform for the whole barn at one time, because of the high number of measures that required a considerable time span. Within this time often animals changed their behaviour. Moreover, lying positions, lying down and rising were best recorded from an elevated

Table 2.1 Overview of farms visited in each country.

		Austria	Germany	Total
Dairy cows	Cubicles	8	8	16
	Deep litter	3	4	7
	Tie stalls	6	6	12
Total		17	18	35
Beef bulls	Deep litter	4	5	9
	Fully slatted floor	5	5	10
Total		9	10	19

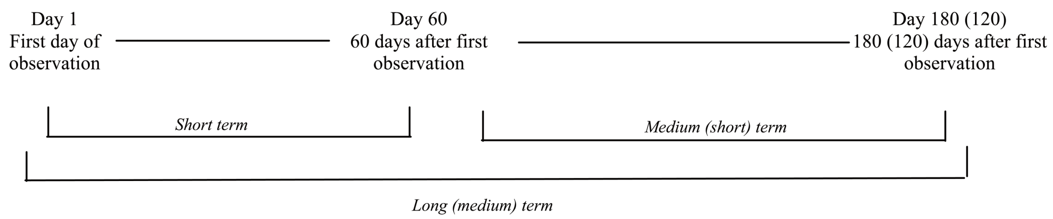

Figure 2.1 Schedule for on-farm observations.

seat, but the high chair often did not allow an overview over the whole pen and needed to be moved from segment to segment.

For fattening bulls, three weight classes were defined in line with the literature and common farming practice:

- Initial fattening period: 200–350 kg (no animals with <200 kg);
- Medium fattening period: >350–550 kg;
- Final fattening period (finishing bulls): >550 kg.

All present weight classes were observed at equal share within each observation hour. Only pens with more than 3 animals were included.

Observation sessions lasted in total about 5 hours including recording of rising and lying down, scan sampling and moving between segments. They started after the main feeding period in the morning, and were carried out by one observer positioned on the feeding table on an elevated observation chair.

2.3.3 Rising and Lying Down

During the development of parameters, we had considered to record durations of rising and lying down events separately for the carpal and remaining phase (for lying down the carpal phase is defined as starting when one of the front legs is being bent, before touching the ground, and ends when the hind quarters start falling down. The remaining phase ends when the animal has pulled the front legs out from underneath the body. For rising it begins when the animal starts lifting the hind quarters from the ground and ends when the first claw (front leg) touches ground. The remaining phase ends when both front legs (claws) touch ground). However, this turned out to be less feasible than recording the total duration (carpal + remaining phase). At the same time, durations of carpal phase and total durations were highly correlated both for rising (observer A: r=0.91; B: r=0.94; n=21) and lying down (A: r=0.82; B: r=0.81; n=30). Therefore, it was decided to only record total durations (definition see Table 2.2).

During the on-farm observations, as far as possible, durations of each voluntary normal rising and lying down occurrence were taken by stop watch. It was intended to record 10 occurrences each for bulls per weight class and 20 each for dairy cows. Interrupted or abnormal lying down or rising movements were only counted. Only medium weight and finishing bulls were observed, because it was expected that durations of rising and lying down in young animals are not yet critical in welfare terms. Moreover, we wanted to exclude a carry-over effect from the rearing conditions regarding possible abnormalities in rising or lying down.

TABLE 2.2 Observed behaviours around resting, their definitions and way of calculation.

Parameter name		Description	Measure
Lying down/ rising	Rising	Event starts when the animal starts lifting the hind quarter from the ground. The rising sequence ends when both front legs touch ground and the animal stands with its whole body weight on all four legs again.	Duration in seconds
	Lying down	Event of lying down starts when one carpal joint of the animal is bent (before touching the ground). The whole lying down sequence ends when the hind quarter of the animal has fallen down and the animal has pulled the front legs out from underneath the body.	
	Collision	During rising or lying down the animal hits against housing equipment with any part of the body.	Ratio of specific events/all lying down or rising events
	Slipping	During rising or lying down at least one claw or leg is accidentally sliding abruptly out of place.	
	Interrupted	The sequence of lying down or rising is not finished by the animal.	
	Horse-like rising	Animal gets up with its outstretched front legs first.	
	Hind quarter first	Animal lies down with its hind legs first and bends the front legs afterwards.	
Lying and other behaviour around resting	Outside lying area	Animal lies with its whole body outside the supposed lying area (cubicle or littered area) – only for dairy cows.	Ratio of number of cows in specific lying position/all cows lying
	Hind quarter out of lying area	Animal lies with hind quarter (both hind legs) outside the lying area (cubicle or littered area) – only for dairy cows.	
	Hind quarter on edge	Animal lies with a considerable part of its hind quarter on edge of cubicle or lying area (in sloped straw flow system) – only for dairy cows.	
	Head resting	Animal is lying with its head positioned in a relaxed way either on the floor, housing equipment or its own body.	
	Hind leg stretched	Animal lies with at least one of its hind legs stretched away from its body at an angle of ≥ 90.	
	Lying on side	Animal lies in lateral position with whole body weight put on one side and legs not underneath the body, either stretched or bent.	
	Backwards	Animal lies backwards in the cubicle with head at the position where the hind quarter is supposed to be – only for dairy cows and cubicle systems.	

TABLE 2.2 CONT. Observed behaviours around resting, their definitions and way of calculation.

Parameter name		Description	Measure
Lying and other behaviour around resting	Sitting	Animal sits dog-like on its hind quarter with front legs fully stretched. Sitting animals are generally included into 'lying' for the calculation of the measures	Ratio of number of sitting animals/all animals lying
	Standing on lying area	Animal is standing in cubicle or on littered area (2 floor system) with at least two legs - only for dairy cows and loose housing.	Ratio of number of cows standing on lying area/all cows on lying area (lying and standing)
	Ruminating	Animal is lying and ruminating which can be recognised by the regular movement of the jaw bones and ears.	Ratio of number of lying animals ruminating/all animals lying
	Synchrony of lying	The maximum proportion of animals lying simultaneously (when different segments are observed, the value is calculated from the total number of segments in order to cover the whole pen).	Maximum number of lying animals/all animals per pen

2.3.4 Lying and Other Behaviour around Resting

Numbers of standing and lying animals in specified lying positions were recorded by instantaneous scan sampling every 10-30 minutes depending on the number of different sections/pens to be observed on the farm so that the whole barn was scanned within one hour or two hours if more than six sections needed to be observed. One hour or 2 hours, respectively, were divided by the number of sections/pens so that within one hour or 2 hours each section was observed once for at least 10 minutes. In case of the whole barn being only one segment, scans were performed every 30 minutes (Table 2.3).

In bulls up to 4 pens per weight class were observed, with each pen being scanned at least twice per observation day.

2.3.5 Decision on Measures

A stepwise approach was taken in order to evaluate the suitability of the measures for an on-farm welfare assessment protocol and the following arbitrary thresholds were set:

1. On-farm incidence had to be higher than 1.0 per hour of farm visit for reliable and feasible recording as well as differentiation between farms.
2. In terms of inter-observer reliability, correlation coefficients between observers for the remaining measures had to be r_s or W>0.70.

TABLE 2.3 Length of scan sampling intervals in minutes for different number of sections/pens.

	Number of sections/ pens											
	Within one hour						Within two hours					
	1	2	3	4	5	6	7	8	9	10	11	12
Scan sampling interval in minutes	30	30	20	15	12	10	17	15	14	12	11	10

3. In terms of intra-farm consistency, again for the remaining measures correlation coefficients between results from different observation days had to be W>0.70.
4. For reliable and consistent measures, results from reduced observation times were checked again for intra-farm consistency.

2.4 RESULTS

2.4.1 Descriptive Measures of Behaviour around Resting

Descriptive measures regarding rising and lying down durations are presented in Table 2.4, and regarding the proportion of different lying positions and standing on lying area in Table 2.5.

The maximum percentage of synchronously lying dairy cows and bulls ranged from 25 % to 100 % with a mean of 62 % (bulls: 61 %) (SD dairy: 12.2; bulls: 13.6).

2.4.2 Incidences of Behavioural Parameters around Resting

Rising and Lying Down

In beef bulls it was often not possible to record the intended number of rising and lying down occurrences within 4 or 5 hours. For dairy cows the same was true for rising (Table 2.6 with occurrences per observation hour).

During lying down movements (903 in bulls, 2,143 in dairy cows), collisions with equipment occurred more often than slipping, which was very rarely observed (Table 2.6).

Overall incidences of events during lying down/rising movements were of the same magnitude in both countries, except for horse-like rising which was observed far more

TABLE 2.4 Durations of rising and lying down in dairy cattle and beef bulls in different housing systems (Mean; Minimum value, Min; maximum value, Max; standard deviation, SD; observed sequences, N).

	System	Rising (seconds)					Lying down (seconds)				
		N	Mean	Min	Max	SD	N	Mean	Min	Max	SD
Dairy cows	Cubicles	882	4.71	1.81	80.00	4.54	1042	5.45	2.28	36.50	2.28
	Deep litter	249	3.50	2.04	8.86	.81	474	4.15	2.25	8.89	1.00
	Tie stalls	427	4.90	1.47	50.78	4.05	627	6.05	2.31	146.16	6.08
Beef Bulls (m)	Deep litter	154	2.99	1.96	4.68	.60	252	3.98	2.02	10.46	1.10
(f)	Deep litter	148	3.20	1.71	5.49	.70	247	4.11	2.10	8.46	1.09
Beef Bulls (m)	Fully slatted	155	3.32	1.39	6.83	.91	250	5.50	2.76	14.61	1.96
(f)	Fully slatted	112	3.65	1.39	7.39	1.12	154	5.93	2.28	15.87	2.10

Notes: m=medium fattening period; f=final fattening period.

TABLE 2.5 Mean percentages of animals lying in different positions or standing on lying area in dairy and beef (all weight classes merged)

	System	Hind leg stretched	HQ not visible	Head resting	Lying on side	Rumi-nating	Sitting	Back-wards	Stan-ding on lying area	Lying partly or completely outside lying area
Dairy cows	Cubicles	16.14	26.99	09/10/24	.67	47/04/01	00/03/01	.00	1.55	1.93
	Deep litter	9.98	48.89	09/09/16	.52	54.70	.00	n.a.	8.75	.00
	Tie stalls	22.60	.28	9.97	.72	47.95	.00		n.a.	48.46
Dairy cows		16.74	24/03/01	09/10/04	.66	48.27	12.08	00/02/01	.00	09/10/22
Beef bulls	Deep litter	5.52	49.96	11.59	09/02/21	4.24	.00	n.a.	n.a.	n.a.
	Fully slatted	8.99	15.58	11.59	1.32	43.85	.00	n.a.	n.a.	n.a.
Beef bulls		6.57	39.59	1.89	1.94	41.33	.01	n.a.	n.a.	n.a.
Total mean		13/04/01	29.68	1.35	09/01/13	45.75	00/02/01			

Notes: HQ=Hind quarter; n.a =not applicable; m=medium fattening period; f=final fattening period; lying partly or completely outside lying area = HQ on edge + HQ out + outside lying area.

often in Austrian dairy cows (81 of 98 sequences observed in total, most of them in cubicle and tie stall systems).

Only rising and lying down as well as collisions in dairy cows occurred at average incidences that allow to reliably record the behaviours within a limited observation time. All other behaviours (slipping, abnormal rising or lying down sequences, and collisions in beef bulls, Table 2.6) were, therefore, excluded from the further analysis.

Lying and Other Behaviour around Resting

The overall incidences of resting behaviours were of the same magnitude in both countries.

TABLE 2.6 Mean frequencies of rising and lying down occurrences per farm visit and hour in dairy cows and beef bulls (medium and final weight classes merged).

			Normal sequences		Abnormal sequences				
		System	Total	Collisions	Slipping	Total	Interrupted	Horselike	Hind quarter first
Rising	Beef bulls	Deep litter	09/01/29	00/01/01	.00	00/09/01	00/01/01	00/08/01	.00
		Fully slatted	.89	00/02/01	.14	.28	00/02/01	.23	.00
	Beef bulls		09/01/08	00/02/01	00/08/01	.19	00/01/01	.16	.00
	Dairy cows	Cubicle	09/04/09	09/03/24	00/03/01	00/09/01	00/02/01	00/07/01	.00
		Deep litter	5.66	.00	.00	00/04/01	00/04/01	.00	.00
		Tie stall	2.96	.15	00/07/01	.34	00/03/01	.31	.00
	Dairy cows		3.78	09/01/15	00/04/01	.21	00/03/01	.18	.00
Mean of rising			09/02/15	.47	00/06/01	.20	00/02/01	.17	.00
Lying down	Beef bulls	Deep litter	09/02/11	.13	.00	00/05/01	00/03/01	.00	00/02/01
		Fully slatted	09/01/26	.35	00/06/01	.39	00/07/01	.00	.29
	Beef bulls		1.68	.24	00/03/01	.22	00/05/01	.00	.16
	Dairy cows	Cubicle	7.00	09/05/14	00/01/01	.13	00/09/01	.00	00/01/01
		Deep litter	8.89	00/08/01	.00	.00	.00	.00	.00
		Tie stall	4.40	1.48	00/03/01	.27	.19	.00	00/09/01
	Dairy cows		09/06/02	2.47	00/02/01	.18	00/12/01	.00	00/04/01
Mean of lying down			3.39	09/01/12	00/03/01	.21	00/08/01	.00	00/11/01

Notes: exceed threshold of 1.0 occurrences/hour.

On average 5 dairy cows were observed standing on the lying area (in cubicles or on littered area) per farm visit and hour. Those frequencies of the other behaviours around resting that occurred more than once/hour are displayed in bold in Table 2.7.

In conclusion, resting measures that are feasible for on-farm welfare assessment behaviour are presented in Box 2.1.

2.4.3 Inter-observer Reliability

Rising and Lying Down

Inter-observer agreement for the total durations of rising and lying down was r_s>0.80 throughout the different test sessions with video and on-farm recordings (Table 2.8).

Slipping during rising and lying down sequences was never observed in any of the IORT sessions. Also, testing IOR for abnormal rising or lying down was not possible as it neither

TABLE 2.7 Mean number of lying animals showing different behaviour around resting per farm visit and hour in dairy and beef cattle.

Number of animals	Lying animals	Rumi-nating	Head resting	Hind leg stretched	Lying on side	Sitting	Backwards	Lying partly or completely outside lying area
Beef Bulls	29.13	12.04*	3.17*	1.91*	.56	.00	n.a.	n.a.
Dairy cows	27.69	13.37*	2.78*	4.64*	.18	.01	.00	3.32*

Notes: * exceed threshold of 1.0 occurrences/hour; n.a. = not applicable; all weight classes merged.

BOX 2.1 Resting measures that are feasible for on-farm welfare assessment behaviour.

Dairy cows	*Beef bulls*
Duration of rising	Duration of rising
Duration of lying down	Duration of lying down
Percentage of collisions during rising or lying down	Percentage of animals lying with hind leg stretched
Percentage of animals lying with hind leg stretched	Percentage of animals lying with head resting
Percentage of animals lying partly or completely outside lying area	Percentage of animals ruminating during lying
Percentage of animals lying with head resting	Synchrony of lying
Percentage of animals ruminating during lying	
Percentage of animals standing on lying area	
Synchrony of lying	

TABLE 2.8 Inter-observer reliability (Spearman correlation coefficient, rS) between observers A and B for total durations of lying down and rising in different test sessions.

	Test session	n	r_s	Test conditions
Duration of lying down	1	34	*.85***	Video, dairy cows
	2	13	*.98***	On-farm, dairy cows
	3	21	*.98***	On-farm, dairy cows
	4	65	*.95***	Video, bulls and dairy
Duration of rising	1	21	*.85***	On-farm, dairy cows
	2	6	*.83***	On-farm, dairy cows
	3	8	*.83***	On-farm, dairy cows

Notes: figures in italics exceed threshold of r_s=0.70.

occurred during an IORT on farm, nor was it possible to record respective videos for IOR testing.

Collisions with housing equipment during rising or lying down were observed during sessions 3 and 4 with very low agreement between the two observers (PABAK=0.20) and moderate agreement in video observations after revision of the definition from ‘hit’ to ‘forceful hit’ (PABAK=0.78). Thus, it appears that based on a more precise definition and possibly after better training sufficient agreement can be reached.

TABLE 2.9 Inter-observer reliability (Spearman rank correlation coefficients, rs, and Kendall's coefficient of concordance W) for lying and other behaviours around resting in different test sessions for observers A, B and C.

IOR session	Analysis	n	Lying[1]	Ruminating	Standing on lying area	Head resting	Hind legs stretched	Lying partly or completely outside lying area	Sitting, lying on side, backwards, hindquarters out
2 on-farm	r_s (A, B)	20	*0.99***	*0.97***	*0.99***	*0.99***	*1.00***		
3 on-farm	W (A, B, C)[2]	30	*1.00***	*0.86**	*0.82**	*0.83**	0.50	not seen	not seen
	r_s (A, B)	30	*1.00***	*0.93***	*1.00***	0.67*	*0.81***		
4 pictures	r_s (A, B)	57	*1.00***	n.a.	not seen	*0.95***	*0.82***	*0.75***	*1.00***

Notes: * p=0.05; ** p=0.01; [1] basis for calculation of synchrony of lying; figures in italics exceed threshold of r_s=0.70; [2] C = less trained; n.a. = not applicable.

Lying and Other Behaviour around Resting

Assessment of lying or standing (on lying area) as well as ruminating while lying was generally highly repeatable between observers throughout the test sessions (Table 2.11). Reliability of the assessment of hind leg positions relative to the rest of the body, or of the head position was less consistent. However, correlation coefficients were all above 0.80, except for 'hind legs stretched' when a third, less well trained observer participated in the testing (session 3, Table 2.9).

Sitting, lying on side, lying backwards in cubicles and with hind quarters outside lying area never occurred during on-farm IORT. However, because the positions are rather conspicuous, they showed a high repeatability when pictures were used (session 4, Table 2.9).

In conclusion, resting measures that can *reliably* be recorded within on-farm welfare assessment are presented in Box 2.2.

2.4.4 CONSISTENCY OVER TIME

Rising and Lying Down

Kendall's correlation coefficients for durations of lying down and rising at the three observation days as well as for percentages of collisions during lying down were all above the threshold of 0.7 in dairy cows (Table 2.10).

Box 2.2 Resting measures that can reliably be recorded within an on-farm welfare assessment.

Dairy cows	*Beef bulls*
Duration of rising	Duration of rising
Duration of lying down	Duration of lying down
(Percentage of collisions during rising or lying down)	Percentage of animals lying with hind leg stretched
Percentage of animals lying with hind leg stretched	Percentage of animals lying with head resting
Percentage of animals lying partly or completely outside lying area	Percentage of animals ruminating during lying
Percentage of animals lying with head resting	Synchrony of lying
Percentage of animals ruminating during lying	
Percentage of animals standing on lying area	
Synchrony of lying	

Table 2.10 Consistency over time (across days 1, 60 and 180) concerning durations of rising and lying down, and concerning the percentage of collisions during lying down.

	Dairy cows		Beef bulls (m & f)	
	Kendall's W	p	Kendall's W	p
Duration of rising	.74	.000	.50	.075
Duration of lying down	.78	.000	*.82*	.001
% collisions during lying down	.95	.000	*n.r.*	n.r.

Notes: * exceed threshold W=0.70; m & f = all data from medium and final weight classes merged; n.r. = not recommended.

When a reduction of observation time was simulated by only taken those occurrences into account that could on average have been recorded within one to four hours (basis for calculations), for dairy cows consistency sharply decreased for durations of rising, but not for lying down (Table 2.11). In beef bulls only lying down observed over 2 hours approached the threshold. However, looking at all results of simulated reductions of observation time in beef bulls, and considering that only two occurrences of lying down per farm were the basis for the farm values, this result is likely to be accidental.

Although for collisions during rising we found relatively high correlations between results from different days, we excluded the measure from further analysis, because it can only be recorded if rising is specifically observed. However, levels of duration of rising showed low consistency for a limited number of occurrences which led to the exclusion of this measure from our candidate list.

For durations of lying down and collisions during lying down in dairy cows, there were good correlations between results from longer and shorter observations (Table 2.12).

Lying and Other Behaviour around Resting

Regarding resting behaviour in beef bulls, only the percentages of lying bulls ruminating showed a sufficient consistency over time in finishing bulls (Table 2.13). In the medium

TABLE 2.11 Consistency over time (across days 1, 60 and 180) concerning durations of rising and lying down, and collisions in dairy cows with simulated reductions of observation time to 1 to 4 hours.

			1 hour	2 hours	3 hours	4 hours
	Lying down	Kendall's W	*.72*	*.74*	*.76*	*.78*
		p	.000	.000	.000	.000
		Number of events	6	12	18	24
	Rising	Kendall's W	.39	.56	.67	.68
		p	.219	.008	.000	.000
		Number of events	3	6	9	12
	Collisions during lying down	Kendall's W	*.88*	*.93*	*.94*	*.94*
		p	.000	.000	.000	.000
		Number of events	6	12	18	24
	Collisions during rising	Kendall's W	*.83*	*.86*	*.86*	*.87*
		p	.000	.000	.000	.000
		Number of events	3	6	9	12
Beef bulls	Lying down	Kendall's W	.50	*.70*	.59	.60
		p	.082	.004	.021	.021
		Number of events	1	2	3	4
	Rising	Kendall's W	.48	.38	.39	.37
		p	.105	.300	.284	.330
		Number of events	1	2	3	4

Notes: figures in italics exceed threshold W=0.70.

TABLE 2.12 Correlations (Spearman r_s) between durations of lying down and collisions recorded during one to four hours of observation (1–4, simulated data on the basis of results from Table 2.8) and total number of recorded occurrences (All).

			1-to-All	2-to-All	3-to-All	4-to-All
Dairy cows	Lying down	r_s	.88	.93	.99	1.00
		p	.000	.000	.000	.000
		Number of events	6	12	18	24
	Collisions during lying down	r_s	.97	.99	.99	1.00
		p	.000	.000	.000	.000
		Number of events	6	12	18	24

TABLE 2.13 Consistency over time (across days 1, 60 and 180) concerning lying and other behaviour around resting in the different weight classes of beef bulls.

		Initial period		Medium period		Final period	
		Kendall's W	p	Kendall's W	p	Kendall's W	p
Behaviour around resting	Hind leg stretched	.21	.870	.43	.177	.63	.207
	Head resting	.37	.326	.57	.031	.41	.225
	Ruminating	.58	.026	.66	.008	*.82*	.043
	Synchrony of lying	.23	.828	.39	.275	.39	.272

Notes: figures in italics exceeds threshold W=0.70.

TABLE 2.14 Consistency over time (across days 1, 60 and 180) concerning lying and other behaviour around resting in dairy cows and beef bulls (medium and final weight class merged).

		Dairy cows		Beef bulls	
		Kendall's W	p	Kendall's W	p
Behaviour around resting	Hind leg stretched	.63	.001	.49	.094
	Lying partly or completely outside lying area	*.87*	.000	n.a.	
	Head resting	.60	.003	.61	.016
	Ruminating	.57	.006	*.74*	.002
	Synchrony of lying	.54	.012	.37	.339
	Standing on lying area	.59	.019	n.a.	

Notes: figures in italics exceed threshold W=0.70; n.a. = not applicable.

weight class, the correlation coefficient was near to the threshold. Finishing bulls are not continuously present at the farms, whereas medium weight bulls are the class that can most consistently be found. Moreover, differentiation between medium and final weight is often difficult. We, therefore also investigated correlations of results over time for the two weight classes merged, but no improvement of consistency was reached by this (Table 2.14). However, as the correlation between percentages of ruminating lying bulls over time was still above a Kendall's W of 0.70, we regard it acceptable for reasons of feasibility to observe both weight classes without differentiation.

In dairy cows, the only consistent measure over time was the percentage of animals lying partly or completely outside the lying area (Table 2.14).

When simulating a reduction of observation time by only taking results from the first or the last two observation hours into account, consistency even slightly increased when the first two hours were selected, but decreased for the last two hours of observation (Table 2.15).

In conclusion, resting measures recommended for an on-farm welfare assessment are presented in Box 2.2.

2.5 DISCUSSION

2.5.1 RISING AND LYING-DOWN MOVEMENTS

Inter-observer reliability for lying down and rising was consistently high in all test sessions (r_s: 0.83-0.98), even with limited experience of the observers in the first session. This was probably due to an intensive training beforehand. In dairy cattle, it was possible to record

TABLE 2.15 Consistency over time (across days 1, 60 and 180) concerning 'percentage of dairy cows lying with hind quarter on edge of lying area' and 'percentage of lying and ruminating beef bulls' (medium and final weight class) when a reduction of observation time to 2 hours (first and second, third and fourth observation hour) is simulated.

Measure	hour 1+2		hour 3+4	
	Kendall's W	p	Kendall's W	p
Lying partly or completely outside lying area (dairy cows)	.88	.000	.80[1]	.000
Ruminating (beef bulls, m & f)	.77	.001	.63	.012

Notes: [1] value of hour 3+4 missing for two farms; m & f = medium and final weight classes merged.

BOX 2.2 Resting resting measures *recommended* for an on-farm welfare assessment.

Dairy cows	*Beef bulls*
Duration of lying down	(Duration of lying down)
(Percentage of collisions during lying down)	Percentage of animals ruminating
Percentage of animals lying partly or completely outside lying area	

reasonable numbers of 3 or 6 events per hour of observation, respectively. In general, behaviours preceding the lying down event such as stepping onto the lying area and sniffing the ground make it easier to observe lying down compared to rising which often starts without conspicuous signs beforehand and thus is easily missed during on-farm observations.

Considering further the aspects of consistency over time and the limited observation available in an on-farm welfare assessment situation, in dairy cows only 'duration of lying down' can clearly be recommended for the monitoring system. Even with only 6 lying down occurrences per farm a sufficiently representative value for the farm was achieved. For rising, our data suggest that not even from an average of 12 rising occurrences per farm a representative value can be calculated for the individual farm. Consistency over time was only acceptable when about 18 to 19 occurrences per farm were available. It may be considered to include 'duration of rising', if it is possible to record the behaviour during a longer time span in which further data are collected. However, at present we do not recommend the measure for feasibility reasons.

In beef bulls, in general much lower numbers of rising and lying down could be recorded. Here the measure 'duration of lying down' can only be recommended, if the welfare assessment protocol allows the recording over a longer time span during which further data are collected. With our data, an acceptable consistency over time was only reached with on average about 8 lying down occurrences per farm (within about 5 observation hours).

Collisions with housing equipment during lying down were rather frequent in dairy cattle. Additionally, consistency was very high in dairy cows (W=0.95). Although inter-observer reliability was partly problematic (PABAK=0.20 and 0.78), after revision of the definition (e.g. 'hit' to 'forceful hit') and intensive training this measure appears to suitable for inclusion into the assessment protocol. However, an improvement in inter-observer reliability would have to be demonstrated in the next stages of the project.

Animals slipping during rising or lying down events as well as all abnormal behaviours around rising and lying down such as interrupted movements, animals lying down with their hind quarter first, animals rising with front legs first (horse-like) were observed very infrequently in dairy and beef cattle (abnormal behaviours: 0.04–0.39 events per hour) during on-farm observations. Therefore IOR testing could not be performed. All these measures were excluded from further analysis and cannot be recommended for on-farm application.

2.5.2 Lying and Other Behaviour around Resting

Animals lying were almost identically recorded by different observers (r_s: 0.99–1.00) and also animals ruminating (during lying) were assessed highly reliably (r_s: 0.86–0.97).

Levels of synchrony of lying (maximum value of animals lying simultaneously) were not consistent over time. However, the method applied in this study of scanning segments instead of the whole barn was not optimal for this measure. Where many segments were observed, in the worst case, it only produced 2 complete scans for the whole group within four hours of observation. This method was chosen because of the high number of measures that had to be recorded at one time. If only a very limited number of measures have to be recorded, it would be feasible and recommendable to scan whole pens –and not segments- at relatively short intervals (10–20 minutes). Then synchrony of lying could tentatively be included and evaluated again as it is easily assessed.

Consistency over time was only sufficient in dairy cows for the ‘percentage of cows lying partly or completely outside lying area’ (W=0.88 for data from first two hours after morning feeding) Inter-observer reliability (r_s: 0.75) could only be tested in one session using photographs (behaviour shown in 6 out of 57 animals). Disagreement might have been caused by the two-dimensionality and the perspective of the pictures. From on-farm observations improved inter-observer reliability can therefore be expected. The importance of training was also shown for the measure lying with hind leg stretched for which the overall agreement between three observers was substantially reduced by one observer who was not as experienced as the others.

In beef bulls consistency over time was acceptable for the ‘percentage of bulls ruminating during lying’ (W=0.77 for data from first two hours after morning feeding and for bulls heavier than 350 kg). However, this measure is very likely heavily influenced by the total number of ruminating bulls. Differences are, therefore, possibly more feeding related than related to resting comfort and not necessarily linked with welfare concerns with respect to resting. We realised that a better option had been to record the ratio of bulls ruminating during lying to all bulls ruminating in order to only capture the extent of rumination associated with lying. However, as the number of standing bulls ruminating had not been recorded, we could not test this measure. We recommend to use further stages of the project to test and evaluate this measure.

All other measures around resting occurred too infrequently during farm visits in both animal categories and therefore can not be recommended.

2.6 CONCLUSIONS

In conclusion, we recommend to include into an on-farm welfare monitoring system for dairy cows the measures 'duration of lying down', taking the total duration of at least 6 voluntary occurrences, and the 'percentage of collisions during lying down' during these occurrences. Additionally, during the first two hours after the morning feeding, the 'percentage of cows lying partly or completely ouside lying area' should be recorded by instantaneous scan sampling over the whole pen every 10 to 20 minutes. During the two hours of observations, further behavioural measures can be recorded.

In beef bulls, we recommend to record the 'duration of lying down', but a minimum of eight voluntary occurrences in bulls heavier than 350 kg should be achieved which might take on average five hours. However, during this time further data can be collected. In further stages of the project possibly the measure 'ratio of lying bulls ruminating to all bulls ruminating' should be tested and evaluated.

3

CLEANLINESS SCORING FOR DAIRY AND BEEF CATTLE AND VEAL CALVES

K.A Leach, U. Knierim and H.R. Whay

3.1 SUMMARY

Cleanliness of cattle reflects the environment in which they are kept and has implications for a number of health issues. There is some evidence that cattle find dirty environments aversive, especially when first introduced to them. Associations between dirty udders and mastitis, and dirty legs and digital dermatitis, have been shown. Dirt on the body may cause skin irritation, although published evidence of this has not been found. A number of systems of cleanliness scoring have been published, for use in scientific studies, or giving management advice.

We propose a simplification of one of these published systems for dairy cows, which reduces the number of categories from four to two – acceptable or too dirty. This would involve scoring dairy cows for cleanliness in three different anatomical areas: udder, lower legs, and hind quarters; since dirt in these areas has different health implications. The proportion of animals which are scored as 'too dirty', according to a set of descriptive criteria, in each of these regions will be recorded. Both milking and dry cows should be assessed, since intra-mammary infections often originate from cows being kept in dirty conditions in the dry period.

Since beef cattle will be more difficult to view safely and satisfactorily, we propose a simpler system, in which the side view of the animal is assessed as a whole, rather than scoring separate areas. Veal calves could be assessed in the same way as dairy cows if individually housed, but if they are group housed this may be difficult.

Experience has shown that hygiene scoring is quite subjective, and training should include clear instructions, discussion, and viewing live animals.

3.2 INTRODUCTION

Cattle are frequently exposed to large amounts of mud and faeces in their environment. There is some evidence from behavioural choice experiments that cattle find these dirty environments aversive, particularly when first introduced to them, and will avoid them if possible (Phillips and Morris, 2002). Such conditions could therefore be considered as compromising the welfare of cattle. Additionally, the hygiene of the environment is known to be related to infectious diseases such as mastitis (Schukken et al., 1990) and digital dermatitis (Roderiguez-Lainz et al., 1996). The cleanliness of the teats and lower limbs is directly important for health, since if these areas are dirty, they are more likely to be contaminated with micro-organisms which may invade the udder, or affect the skin. A direct relationship with subclinical mastitis has been reported by Schreiner and Ruegg (2003). The presence of mud or faeces on the skin may be irritant, causing discomfort. The cleanliness of animals themselves reflects the nature of their past environment, even if very recent measures have been taken to disguise poor environmental hygiene, and therefore can be a useful indicator on farm welfare visits.

Assigning scores for different areas of the animal gives indications as to the source of contamination. For example, dirty lower legs imply cows are walking through mud or slurry, dirty flanks indicate dirty and wet lying conditions. A particular pattern of manure splashed on the flanks is a sign that the cow has been flicking its tail, covered with thin faeces. This combination of behaviour and faecal consistency may be associated with disturbed rumen function (Huxley and Whay, 2006).

In young calves, dirty flanks and perinaeal area will reflect the presence of gastrointestinal problems – if the calves have suffered from or are suffering from scours, the flanks and perinaeal area are likely to be more dirty.

The cleanliness of cattle has been introduced relatively recently as an indicator in scientific and veterinary investigations relating to animal health and welfare. A small number of similar scoring systems, referred to as ‘cleanliness scores’ or ‘hygiene scores’ has been developed.

3.3 METHODS

Searches were made of the scientific literature and internet resources to discover cleanliness/hygiene scoring systems which have been used. The feasibility of using the various systems was considered, as was the relevance of the assessment to different types of cattle. Repeatability of the proposed systems for dairy and beef cattle was tested between two experienced observers assessing photographs. An inexperienced trainee's scores were also compared.

3.4 RESULTS AND DISCUSSION

The first published system of assessing cleanliness of cows was proposed by Faye and Barnouin (1985). Five areas: the anogenital area, back of the udder, the bottom part of the hind legs (from the hock to the dewclaws), the udder sides and belly and the thighs were clearly defined. The cleanliness of each area was scored on a scale, where:

0 = no dirt;
0.5 = some small dirty parts;
1 = large dirty parts covering less than half the area;
1.5 = large dirty parts covering more than half the area;
2 = area completely covered with dirt.

The scores were illustrated by drawings. When an area had two sides, only the dirtier side was scored. This system has been used in welfare assessment on the continent, e.g. by Krebs et al. (2001). It was used in a modified form by Napolitano et al. (2005). The scores for all five regions are often summed to give a single value for the animal.

Bergsten and Pettersen (1992) assessed cleanliness when investigating the effects of electronic position training for cows in tie-stalls. They assessed five areas of the body: each side above and below the hock, viewed laterally, and the tail end directly from behind, from video films, and awarded scores from 0 (perfectly clean) to 3 (very dirty). Two observers scored independently, and then an iterative process was continued until a score was agreed upon. However, publications gave no clear description of these categories.

Hughes (2001) developed a scoring system for use in the dairy industry, which has been applied, for example, in observations on the association between cleanliness and mastitis (Ward et al., 2002). Separate scores, on a scale from 1 (very clean) to 5 (heavily soiled), are allocated to the flanks, hind legs, udder and tail. A series of photographs is given to illustrate the scores. The cow is scored from both sides, and if scores differ, the higher score is taken.

Another system has been developed and applied for practical use on farms by Cook, at the University of Wisconsin-Madison School of Veterinary Medicine, which can be found at: <http://www.vetmed.wisc.edu/dms/fapm/fapmtools/hygiene.htm>.

This categorises manure contamination on the udder, lower rear limb, and upper limb and flank, with verbal, diagrammatic and photographic descriptions of four categories for each site. The system has been widely used in extension work.

Schreiner and Ruegg (2002) used leg and udder hygiene scores when investigating the influence of tail docking on cleanliness and subclinical mastitis. They used model photographs to illustrate four categories:

1. completely free of dirt or has very little dirt;
2. slightly dirty;
3. mostly covered in dirt;
4. completely covered, caked in dirt.

An alternative very simple method is used in the UK 'BWAP' welfare assessment scheme (Leeb et al., 2004). The same areas – lower limb, udder, upper leg and flank are scored as clean or dirty, with the threshold for dirtiness being a patch of dirt the size of a hand.

In the UK, a cleanliness scoring system for cattle at slaughter is in regular commercial use. In the mid 1990's, following outbreaks of E. coli 0157 in the UK, abattoirs introduced UK Standards for Cleanliness of Cattle Submitted for Slaughter. This system is in place for food hygiene reasons, but it could equally be used in welfare assessments. Although the assessment concentrates on areas of the body important in the slaughter process, excessive dirt in these areas suggests the animal has been kept in unhygienic conditions generally. There are 5 categories, ranging from 1 – clean and dry, to 5 – filthy and wet. The Food Standards Agency (2002) publication 'Red Meat Safety and Clean Livestock' gives photographic examples of animals in each category, but no clear verbal description.

3.4.1 Selection of a Cleanliness Scoring System

All the existing systems used for dairy cattle are largely similar, with the main differences being the degree of detail, and format, of the description for each category. The criteria for choice were a simple system, with clear descriptions, concentrating on the clinically important areas.

It is suggested that a simplification of Cook's system should be used (Table 3.1). This covers the clinically important areas for dairy cows, i.e. the lower limb and udder, with additional information on the flanks. The condition of the flanks gives an indication of the cleanliness and dryness of the lying area, which is important in terms of environmental mastitis. We propose including the rear view of the hind quarters (above the udder) also, since dirt from here may transfer to the udder. The terms 'splashing' and 'plaques' need to be more clearly defined, particularly as we propose that scores 3 and 4 indicate the animal is 'too dirty', and therefore the distinction between 2 and 3 is important. It is suggested that 'plaques' should apply to areas of dirt which either cover more than half the area being assessed, or have a 'third dimension', ie a visible thickness to the layer of dirt. Each body area would be classed as 'acceptable' or 'too dirty', and the proportion of

TABLE 3.1 Proposed cleanliness scoring system for dairy cattle and calves.

Region	Acceptable		Too Dirty
	Score 1	Score 2	Score 3
Lower hind legs (coronary band to hock)	Little or no manure above the coronary band	Minor splashing above the coronary band	Separate or continuous plaques of manure above the coronary band
Hind quarters – upper hind leg (above the hock), flank, and rear view excluding udder	No manure present	Minor splashing of manure	Separate or continuous plaques of manure
Udder (not calves)	No manure present	Minor splashing of manure near the teats	Distinct plaques of manure on udder or any dirt on and around the teats

Notes: based on Cook, n.d.

animals sampled scoring 'too dirty' would be the summary measure used. Both sides of the animal should be viewed if possible and the dirtiest side scored. The same system could be used for veal calves.

For beef cattle, we recommend that the assessment should be made from outside the pen, for safety reasons. Therefore, scoring specific areas of a selected sample of animals is not feasible, and a simpler, 'whole animal score' is required. We propose that an assessment is made at group level of the proportion of animals in which more than a quarter of the side view (excluding head and lower legs), is covered with plaques of dirt.

We define 'plaques' as areas of dirt which have a third dimension (visible thickness) or cover more than half the region being assessed.

3.4.2 Repeatability and Reliability

Cleanliness scoring using a five point system, based on similar diagrammatic representations to those of Cook, was tested for between-observer reliability by Faye & Barnouin (1985), by comparing the scores of 9 observers for the various areas of the body with those of an 'expert'. Spearman correlation coefficients ranged from 0.67 to 0.94, depending on the observer and the area being assessed. Divergence was greatest for the lower leg and least for the udder. All correlations were highly significant ($p < 0.001$). Schreiner and Ruegg (2003) analysed the repeatability of scoring within observer, using their four point system, by duplicate scoring of 100 cows. Repeatability was 77% for udders and 85% for legs. When scores were grouped as 'Clean ' (score 1 or 2) and 'dirty' (score 3 or 4), repeatability increased to 95% for udders and 96% for legs.

3.4.3 Validation

Cleanliness is considered very important for dairy cows, in view of the implications for intramammary infection, and the possible indication of rumen dysfunction (animals might be more dirty if faeces were very loose due to rumen upsets). The relationship with risk factors for lameness (prolonged contact with large amounts of slurry) is also important. For beef cattle the main relevance would be in relation to lameness, possible rumen dysfunction, and the general level of care of the animals. For veal calves, the main issue reflected in cleanliness is likely to be the occurrence of scours. An assessment of the flanks of the animal, and the rear view of the perinaeal area, are considered the most relevant for this group, although there is no published literature on the cleanliness of veal calves.

3.4.4 Feasibility

In terms of sampling, Hughes (2001) recommends scoring 20 cows at random to assess a herd, taking approximately 30 minutes, while Cook advocates whole herd scoring in herds under 100 cows, or 25% of each group in larger herds. In the context of a welfare assessment, it would be reasonable to score the same random sample of animals as is selected for condition scoring, and clinical scoring, in the case of a dairy herd. Individually penned veal calves could be sampled in a similar way to dairy cows. Sample sizes (see separate chapter) will be large for many EU dairy herds if 95% confidence intervals and 5% absolute precision are required. For beef cattle, it will be necessary to conduct the assessment from outside the pen, and therefore at group level, for reasons of feasibility and safety. This may pose difficulties in observing the legs, if the animals are lying down. There is a risk of either under estimation, or double sampling of the animals which are too dirty. Therefore we suggest that the assessment is made at group level, using the whole animal assessment as described earlier and recording the proportion of the group which are 'too dirty'.

3.5 CONCLUSIONS

Cleanliness does have a relationship with animal welfare, through links with mastitis, lameness and gastrointestinal problems. A system of cleanliness scoring simplified from that of Cook is proposed for use on dairy cows and veal calves, applying random sampling. The proportion of the animals scored which are unacceptably dirty in a particular body area will be the outcome measure. Three body areas will be assessed, which have implications for different aspects of welfare. For beef cattle, detailed examination of individuals is not feasible and a simple assessment of one side of the body is all that could be expected.

4

CARCASS DAMAGE, DEAD ON ARRIVAL AND DOWNGRADES AT SLAUGHTER IN CATTLE

R. Westin, A. Velarde, A. Dalmau and B. Algers

4.1 SUMMARY

Clean, healthy, unstressed animals and animals free from blemishes are profitable animals both from an economic and welfare point of view. To evaluate such parameters, questionnaires have been sent to 10 abattoirs (4 in Belgium, 3 in Spain, 3 in Sweden) with questions regarding present procedures for measuring and documentation of 'dead on arrival' and animal welfare related parameters resulting in carcass damage and downgrades in the slaughter line.

'Dead on arrival' and other welfare related parameters possible to detect at clinical live animal inspection at slaughter are unsuitable candidates for auditing of animal welfare in cattle since they are rare and documentation records are not available in all abattoirs.

A scoring system developed and tested by Swedish Meat Industry Association (Kött and Charkföretagen, 2006) is concluded to be a promising method for assessment of cleanliness of cattle within Welfare Quality. The measure reflects the farm hygiene conditions. 'Dead on arrival' and other welfare related parameters possible to detect at clinical live animal inspection at slaughter are unsuitable candidates for auditing of animal welfare in cattle since they are rare and documentation records are not available in all abattoirs.

In assessment of cleanliness each animal is individually scored into one of 4 categories depending on the level of dried in manure. Fresh faecal contamination is not recorded.

- Category 0 – 'clean': animals which are clean or slightly soiled with dried in manure in critical areas.
- Category 1 – 'moderately soiled': animals which are noticeably soiled with dried in manure in critical areas.
- Category 2 – 'heavily soiled': animals which are heavily soiled with dried in manure or 'manure armour' in critical areas.

- Category 3 – 'extremely heavily soiled': animals which are extremely heavily soiled with dried in manure and/or 'manure armour' in critical areas, large spread and/or urine/manure burn marks in the skin.

The 'critical areas' monitored are:

- under surface/mid line of abdomen;
- Under surface of the chest;
- hook and Achilles tendon;
- under surface of the neck;
- udder and genitalia;
- area around the anus, including rear part of udder.

4.2 INTRODUCTION

Animals presented for slaughter needs to be clean, healthy, unstressed and free from blemishes in order to be profitable. Dirty animals increase the risk for faecal contamination of the meat, the carcase of a sick animal will be fully or partly rejected, stress and bleedings decrease meat quality and may cause meat rejection (Greogry, 1998). Besides the economic value, these parameters are related to animal welfare. Animals arriving dead or dying during lairage also are considered important both from an economic and a welfare point of view. The aim of this study is to define and standardise measures of 'dead on arrival', and animal welfare related parameters resulting in carcass damage and downgrades in the slaughter line. Measuring of bruising and meat pH is not included in this study.

4.3 METHODS

Questionnaires have been sent to 10 abattoirs (4 in Belgium, 3 in Spain, 3 in Sweden) with questions regarding present procedures for measuring and documentation of 'dead on arrival' and animal welfare related parameters resulting in carcass damage and downgrades in the slaughter line. Visits have been made to three abattoirs when veterinarians and slaughter house personnel were interview and different clinical findings were documented on photo.

4.4 RESULTS

In all abattoirs deaths during transport of cattle were very unusual. In one abattoir it had never occurred. In the other companies the number of ‘dead on arrivals’ varied from 1/9,000 to 1/30 000 slaughtered animals. In two abattoirs ‘dead on arrivals’ were not documented at all. In 7 this was documented on the trade/transport document. One abattoir did not say if or how ‘dead on arrivals’ were documented.

The only animal welfare related parameter that may cause downgrading of the carcase that is assessed by all companies except one was cleanliness of the animal. The 3 Swedish abattoirs use a standardised scale with 4 categories of cleanliness (0–3) devised by the Swedish Meat Industry Association (Kött & Charkföretagen, 2006). This scale is implemented in all large Swedish abattoirs (Rutegård, 2006, personal message). A similar scale is used in one of the other asked abattoirs. This scoring system is based on 5 categories insted of 4 (--, -, 0, +, ++). Five companies only assess the animals as clean or not (0–1).

Other parameters related to animal welfare possible to detect during live animal inspection are clinically assessed by the veterinarians working at the slaughter plants. Clinical findings easy to detect, that may result in immediate rejection or downgrades at the slaughter line are fractures, abscesses, wounds and exhaustion (Askar, 2006, personal message). Overgrown/deformed claws and broken horns or horns growing into the scull are also findings related to animal welfare but they will in most cases not cause any downgrading or meat rejection (Askar, 2006 personal message). In most of the asked abattoirs these findings were not regularly documented during live animal inspection. According to veterinarians working at slaughter plants all these findings nowadays are extremely rare (Askar, 2006 personal message; Brattberg, 2006 personal message).

4.5 DISCUSSION

Deaths during transport and clinical findings possible to assess at live animal inspection related to animal welfare such as fractures, abscesses, wounds, exhaustion, overgrown/deformed claws and broken horns or horns growing into the scull are rare. An observer can therefore not expect to find any of these findings during a short visit to an abattoir. Abscesses, old wounds, overgrown/deformed claws and horns growing into the scull are related to farm conditions wile fractures, broken horns, exhaustion and eventually fresh wounds are more related to transport and lairage conditions. To be able to use these parameters in an audit system all abattoirs would need to record those findings systematically. This is however not done at the moment. ‘Dead on arrival’ is often documented but since it occurs so seldom we conclude it to be an unsuitable candidate for auditing of animal welfare in cattle.

The cleanliness of animals reflects the hygiene level of their surrounding. Clean animals are thus likely to be living under good farm conditions. Cleanliness was assessed to some extent in all asked abattoirs except one. Assessment of cleanliness is possible to conduct during regular slaughter procedures. The method devised by the Swedish Meat Industry Association (Kött & Charkföretagen, 2006) is tested an evaluated with good results (Rutegård, personal message). We therefore conclude that this is a promising method for assessment of cleanliness of cattle within Welfare Quality. Since the measure is an indicator of farm conditions, it should be used in farm assessments of animal welfare even if it is audited at abattoirs.

4.6 CONCLUSION

The scoring system developed and tested by Swedish Meat Industry Association (Kött & Charkföretagen, 2006) is a promising method for assessment of cleanliness of cattle within Welfare Quality. The measure reflects the farm hygiene conditions. 'Dead on arrival' and other welfare related parameters possible to detect at clinical live animal inspection at slaughter are unsuitable candidates for auditing of animal welfare in cattle since they are rare and documentation records are not available in all abattoirs.

5

LAMENESS IN DAIRY AND BEEF CATTLE AND VEAL CALVES

K.A. Leach, C. Winckler and H.R. Whay

5.1 SUMMARY

Lameness constitutes a major welfare issue in cattle, causing pain and alteration to normal behaviour. Recent figures for the prevalence of lameness in dairy cattle in European countries range from 22% (Whay et al., 2003) to 45% (Winckler and Brill, 2004) for loose housing systems, and from under 1% up to 21% for systems in which cows are tied for at least part of the time (Bielfeldt et al., 2005; Sogstad et al., 2005). The prevalence in beef cattle is generally lower.

Scoring of gait by observation while animals are walking remains the most feasible measure of lameness, despite its subjective nature. Examination for lesions is too labour intensive, and automatic methods of measuring floor-foot forces or weight-bearing are not sufficiently developed for regular practical use. However, gait scoring requires that individual animals can be clearly seen, walking on a suitable surface. This is possible with loose housed dairy cows, but cannot be achieved for tied cows, or the majority of beef cattle, which are unaccustomed to handling.

For loose housed dairy cows, we recommend locomotion scoring using a three point scale – 0: sound, 1: lame: 2 severely lame, based on a published, validated scale. The prevalence of lame and severely lame cows in the herd can be calculated. Percentage agreement between two observers was 83% for the proposed scale. For tied cows we have developed a new ‘standing scoring’ system, based on the stance of the cow and movement within the stall. This method gives acceptable repeatability (70% agreement within and 89% between observers), and, although it is less sensitive than locomotion scoring, a correction factor can be applied to adjust the prevalence for comparison with herds assessed by locomotion scoring. Random sampling of the milking portion of dairy herds with over 30 cows is suggested. However, power calculations indicate that a large number of cows will need to be assessed in many EU dairy herds. Therefore the time required will probably be considerable. The animals selected could also be used for assessment of other clinical parameters, although dry cows should also be included in the sample for other parameters.

For beef cattle, due to practical and safety issues associated with handling, the only feasible possibility applicable to all farms and systems would be group assessment to identify the number of severely lame animals detected in a group of known size. However, the sensitivity of detection is likely to be low and influenced by the housing system. Thus assessments in different types of housing could not be fairly compared. There are no studies validating this approach. There are no validated methods for lameness assessment in veal calves either. Lameness could not be easily assessed in individually penned calves, but group housed calves might be assessed in the same way as beef cattle.

5.2 INTRODUCTION

Lameness is a definite welfare problem, indicating pain (Whay et al., 1997) and restricting normal behaviour in cattle (Singh et al., 1993; Hassal et al., 1993). A wide selection of methods for assessing lameness by observing the gait of cattle has developed in recent years (Whay, 2002). These are often referred to as 'locomotion scoring' or 'gait scoring' systems. Some of these are more suited to experimental work, being very detailed, while others are simpler and more appropriate for monitoring at herd level. The presence of clinical foot lesions is another indicator of lameness in cattle, but detecting these lesions requires intensive expert handling of the animals, including lifting their feet, and therefore is not applicable for the purposes of monitoring schemes at herd level. Technology measuring the forces under the feet has been used to demonstrate differences between lame and sound cows (Rushen et al., 2006), and relationships between reaction forces and gait (Dyer et al., 2004) but is not at the stage of development to be feasibly useful in commercial welfare assessments. Gait scoring currently remains the only feasible approach.

5.3 METHODS

A review of lameness assessment methods was made by Whay (2002), which served as the starting point for the possible existing methods of lameness assessment which could be used. The range of systems considered included those developed by Manson and Leaver (1988), Sprecher (1997), Whay et al. (1997), Breuer et al. (2000) and Winckler and Willen (2001), and the system currently used by RSPCA Inspectors when carrying out welfare assessments on farms in the UK: 0 (sound), 1 (abnormal but not lame), 2 (lame), 3 (severely lame).

Once a lameness scoring system had been decided upon, the practicalities and approaches for applying it to dairy cattle, beef cattle and veal calves were discussed.

A new method was developed for assessing lameness in tied cows, since none was available (Leach et al., 2009). With the help of a veterinary surgeon experienced in the husbandry of cows kept in tie-stalls, a number of potential indicators for lameness in standing cows were identified. These were: rotation of a foot from the medial axis of the body, resting a foot, repeated weight-shifting ('stepping'), standing on the edge of a step to relieve pressure on a foot, and reluctance to bear weight on a foot when moving sideways. Four observers together carried out pilot exercises, assessing tied cows using these criteria and then scoring the cows as lame or not lame when walking. Following discussion, the assessment procedure was formalised. Three observers used this formalised procedure to assess 138 tied cows on 7 farms. 110 of these were scored by all three observers, and the rest by only 2 of the observers. On three of these farms, locomotion scoring was also carried out by one observer, giving a score of lame or not lame for 76 cows in total, to allow comparison of 'standing score' and locomotion score. Data for testing intra-observer repeatability of the standing score was gathered on two farms where the same observer scored the same cows (n = 46) 6 to 12 hours apart. Comparison between standing score and a five point locomotion score (Winckler and Willen, 2003) was later carried out by two observers on four farms, with a total of 99 cows. These data sets were used to test repeatability, sensitivity and specificity of the new method, and the relationships between lameness prevalence rates detected using the 'standing score' and locomotion score.

5.4 RESULTS AND DISCUSSION

5.4.1 Selection of a Lameness Scoring Scale for Cows Able to Move Freely

Manson and Leaver's scale (Manson and Leaver, 1988) was not considered suitable as it is complex, and has a strong emphasis on subtle alterations in gait without lameness; also a behavioural component is incorporated which is not clearly defined. The arched back criterion which is crucial to Sprecher's scale (Sprecher, 1997) is often not seen in heifers which are lame (Whay, personal communication); also, this indicator was not affected by the use of anaesthetics in lame cows (Rushen et al., 2006), suggesting it may be a long term learned response to chronic pain, and may persist beyond the duration of lameness. Therefore this system was not considered suitably robust. The 'headbob' included in the system of Breuer et al. (2000) was not considered to be helpful as a primary distinguishing factor, so this system was excluded.

The decision was made to use a scale with three categories.

0 Not lame. Timing of steps and weight-bearing equal on all four feet.
1 Lame – i.e. imperfect temporal rhythm in stride creating a limp (irregular foot fall – uneven temporal rhythm between hoofbeats, weight not borne for equal time on each of the four feet).
2 Severely lame. Strong reluctance to bear weight on one limb, or more than one limb affected.

This amounts to a simplification of the scheme of Winckler and Willen (2001), with categories 1 and 2 combined into 'not lame' and categories 4 and 5 combined into 'severely lame', or a simplification of the RSPCA method, again with 'sound' and 'abnormal but not lame' combined. After considerable discussion, it was decided that separating out the 'abnormal gait' was an unnecessary complication, since lameness per se was the critical issue in this case.

5.4.2 Development of a Lameness Scoring System for Tied Cows

The following indicators for lameness in standing cows were validated: rotation of a foot from the medial axis of the body, resting a foot, repeated weight-shifting ('stepping'), standing on the edge of a step to relieve pressure on a foot, and reluctance to bear weight on a foot when moving sideways. All of these, with the exception of rotation, occurred significantly more often in cows which were lame when walking than in those which were sound. Following analysis of the individual indicators, in relation to locomotion score, it was confirmed that, to optimise sensitivity and specificity, a cow showing a single indicator, with the exception of rotation, should be classified as lame.

5.4.3 Validity and Reliability

Work by Whay et al. (1997) indicates that lame cows, as defined by locomotion scoring, have a reduced nociceptive threshold, and improvement in gait of lame cows has been reported after administration of analgesics (Sedlbauer et al., 2006). These findings provide validation for locomotion scoring as a welfare indicator. The fact that more severe lameness presents a greater overall welfare insult (e.g. greater restriction on movement, reducing the animal's ability and or willingness to reach or compete for resources) was the reason for differentiating two levels of lameness severity.

In terms of reliability, Winckler and Willen's five point system achieved 68% agreement between observers, with observers differing by one point in 30% of cases (Winckler and Willen, 2001). The majority of disagreements occurred between sound cows and those with 'abnormal locomotion' but not lameness. If the 'sound' and 'abnormal' categories are amalgamated, as in the proposed scheme, much of this disagreement is removed and 83% agreement is achieved. The scheme chosen employs familiar concepts, and is simple, with clearly defined categories. It is feasible to carry out in many farm situations.

Agreement between observers on the standing score ranged from 85 to 93 % (PABAK 0.7 to 0.86). Within observer repeatability was 70% for one observer scoring 46 cows twice within 12 hours. The scoring system for tied cows showed 78 to 87% agreement with locomotion score, depending on the observer. Scoring the standing cow was less sensitive than locomotion scoring, with lameness prevalence being underestimated by 25 to 33%, depending on farm and observer. Regression of prevalence of lameness detected using locomotion score (Pls) on that detected using standing score (Pss) using the data for 3 observers on 4 farms revealed the following relationship:

$$Pls = (1.54 * Pss) - 2.3 \; (r^2 = 9.47)$$

This equation, or a simple correction factor of 1.5 could be used to correct for the underestimation caused by the lower sensitivity of the standing score.

Specificity of the standing score for detecting cows which were lame when walking was always over 0.81.

5.4.4 Sampling

There are two possible approaches for lameness scoring – either

1. the whole herd is scored; or
2. a representative (random) sample of animals is scored.

On balance, in terms of feasibility, it was decided that for loose housed dairy herds, the representative sampling method should be used, since the time required to score all animals in large herds would be prohibitive. Power calculations are used to calculate the number of cows that need to be scored to estimate the prevalence of lameness within a group of a certain size (a finite population). A minimum number of cows to score should be set, to ensure that small herds are properly assessed – we suggest 30 cows. In general, in tied herds all the cows should be assessed. These herds are likely to be relatively small, and the test is less sensitive than locomotion scoring.

In many cases it will not be practical, safe or possible to have sufficient control over beef cattle to direct them for standard locomotion scoring. Housing for beef cattle tends to be less well adapted for accessibility and controlled movement of animals than that for dairy cattle, and the animals are less accustomed to handling. In these cases it may only be possible to detect severely lame animals, from observations made at a distance, often from outside the pen. This amounts to group assessment, rather than individual animal assessment. In fact, because the sensitivity of detection even of severely lame animals will be affected by stocking rate and housing design, this method cannot be used as a fair comparison between farms. Any lame animals identified during such observations should be indicated to the farmer, and the assessor should ensure that these animals have been appropriately treated.

For groups of veal calves, individual marking and scoring might be possible, depending on pen size, group size and the demeanour of the animals. For large or flighty groups, the same comments as for beef cattle will apply.

For young individually housed veal calves locomotion scoring is not feasible, and has not been validated at all. Therefore no recommendations can be made at present.

5.4.5 Carrying out the Assessments

In dairy herds in which the cows are free to move, the widely accepted approach for all locomotion scoring systems can be employed for the milking portion of the herd, i.e. an observer watches each cow individually, walking freely on a smooth, hard, non-slippery surface. An ideal time to do this is as cows leave the milking parlour, however, this may not be feasible in large herds where milking may take up to 4 hours. If in a separate group, the dry cows may be more difficult to control in this way, particularly if they are out at grass. Thus we recommend limiting observations to the milking cows. In herds where cows are kept in tie-stalls but regularly released, conventional locomotion scoring can be carried out by releasing the cows individually and observing their gait.

Where cattle are tied and it is not possible to release them to observe locomotion, an alternative approach must be taken, based on behavioural observations and animal measures, which can be made with the cows in the stalls. The cow is observed when standing undisturbed, and when it is moved from side to side in the stall. Using a series of indicators cows can be classified as lame or not lame.

For beef animals, it will be much more difficult to control the movement of individuals, and the only possibility is for group assessment. It is recommended that in this case, each group should be observed from a safe point, and the number of severely lame animals noted. The assessor should attempt to observe all animals in the group walking, if this can be done safely. If a lame animal is noted, the assessor should ascertain from records that the animal has been treated suitably; if not, this should be an immediate requirement. If more than a threshold proportion of severely lame cattle are recorded, this should trigger a more detailed investigation, at a later date. (The threshold level needs to be decided upon at the next stage of the consultation process). The possibility of locomotion scoring beef cattle at loading when they leave the farm, or at the point of slaughter, was considered. However, this would require a separate visit from the main assessment, and was therefore discounted. (If in fact, other parameters are being assessed when the animals arrive for slaughter, locomotion scoring might be possible as they are unloaded). It is recognised that the group assessment method suggested for beef cattle has a high risk of false negatives, dependent on group size and the design of the building. The lameness will also be underestimated if the animals are observed walking on a soft surface, e.g. straw bedding. Also, the sensitivity of the method will depend upon the stocking rate and housing system, and therefore this method cannot be used fairly to make comparisons between farms.

As described in the sampling section, locomotion scoring of veal calves has not been validated. It is not appropriate for individually penned animals, but for group housed veal calves, individual marking and scoring might be possible, depending on pen size, group size and the demeanour of the animals. For large or flighty groups, the same approach as with beef cattle would have to be taken.

5.5 CONCLUSIONS

Lameness is considered an important indicator of welfare in cattle. Locomotion scoring of individual animals is a feasible, validated and well recognised method of assessment for dairy cattle which are able to move freely. A very simple three point scale distinguishing lame and severely lame animals from those which are sound is considered most appropriate for inclusion in welfare assessment. A method for assessing tied cows based on stance and movement within the stall has been developed. This is less sensitive than locomotion scoring but identifies approximately 60% of lame cows and 70% of severely lame cows, as defined by locomotion scoring. A regression equation provides a correction factor which can be used to adjust the prevalence of lameness detected using the standing score to an estimate of the prevalence which would be detected using a locomotion score. There is no validated method for lameness assessment of veal calves. Detection of severely lame animals is the only feasible option with beef cattle and probably for group housed veal calves also. The strong influence of housing conditions on the sensitivity of these observations render them unsuitable for making comparisons between farms. They can only serve to draw attention to some of the most severe problems. A threshold level for severely lame animals must be decided upon, and if this threshold is exceeded, further investigations should be carried out. Likewise, acceptable proportions of the dairy herd in each lameness category need to be defined.

6

SCORING OF CATTLE: INTEGUMENT ALTERATIONS OF DAIRY AND BEEF CATTLE AND VEAL CALVES

H. Schulze Westerath, K.A. Leach, H.R. Whay and U. Knierim

6.1 SUMMARY

Literature was reviewed and personal experiences taken into account in order to propose an assessment system for skin alterations in cattle that can be used in the framework of a broader on-farm welfare assessment tool. In the published studies a wide variety of different methods regarding detail as well as regarding type of the qualitative and quantitative descriptions of the alterations were used. Moreover, there was no scheme covering all alterations of the integument. Instead, usually particular body parts or types of alterations were investigated depending on the focus of the study. There was an almost complete lack of published data on the issues of validity, reliability and feasibility. Only two studies provide data from reliability testing and also on time needed for the assessment. On this basis, we propose a new assessment scheme that includes general aspects from a number of published systems and differentiates between lesions or swellings on the one hand and only loss of hair on the other. As the on-farm assessment will often only allow for visual assessment from a certain distance, only alterations of a minimum diameter of 2 cm at the largest extent and overt swellings can be taken into account. The number of swellings/lesions (i.e. broken skin either in the form of a scab or a wound dermatitis due to ectoparasites and (partly) missing teats) and of hairless patches (hyperkeratosis possible) shall be recorded, assessing one complete side of the animal including udder and prepuce in dairy cows and bulls, but excluding the underside of the belly and inner sides of the legs above the carpal and tarsal joints, as they are often not clearly visible. Four measures of skin alterations are generated from these recordings: the mean number of animals affected on the farm per alteration category and the mean number of alterations per category per animal. The scheme can be applied to dairy cows, beef cattle and calves. The proposed assessment system is easy to learn and simple to apply. We estimate that the assessment takes no longer than 2 minutes per animal depending on the number of alterations, accessibility of the animals and complexity of the farm regarding groups and buildings. Assessors should be trained with the aid of photographs of various affected animals and additionally on farm. Reliability should be tested during training.

6.2 INTRODUCTION

The integument is the skin, including its various layers and their appendages. In the context of a welfare assessment of cattle, the hair, skin and claws are most important. Adjacent tissues such as muscles or joints may also be affected and are typically included in an assessment of the integument. Whether certain parts of the body that belong (in part) to the integument, such as tails or horns, are removed (mutilations) is also welfare relevant, and should be considered in association with an assessment of the farm management. Alteration of claws or joints may result in lameness, but scoring of lameness will be covered in another report. As an assessment of the claws' condition is not feasible within a short-term welfare assessment, only the assessment of hair and skin (including joints) will further be considered.

The scope of our work is to review the literature with respect to methods used for the assessment of alterations of the integument in cattle (dairy, beef and veal calves). Alterations of the integument may be due to different causes, and poor nutrition may, additionally, play a role with regard to hair condition and to a possible predisposition for lesions. The different causes can largely be differentiated by the type of the resulting alteration. For explanatory reasons, a brief overview on the different types of possible alterations will be provided. However, the exact origin of the alteration generally has only little effect on the welfare state of the animal. Therefore, we will focus our further discussion on different published assessment methods on general descriptions of alterations rather than on cause-specific ones. The reason for this is our goal to identify an assessment scheme that is easy and quick to learn and apply, and that has good reliability. One keyword to meet these requirements is simplicity. If the assessment scheme were required to lead to the provision of advice on how to improve conditions, more differentiation according to possible causes would be necessary.

6.2.1 Types of Alterations of the Integument and their Specific Assessment

Alterations Due to External Parasites

Infestations with the ectoparasites described below are enhanced by malnutrition and also by humid housing conditions resulting in higher numbers of more severely infested animals in indoor housing especially in winter. Young animals and animals with long hair are more likely to be affected by ectoparasites (Rosenberger, 1970). In cattle, two types of skin disease due to ectoparasites are common:

Mange (scabies)
Mange is caused by mites. The mites either burrow and feed on dermal structures (sarcoptic mange) or live at the skin surface and feed on epidermal debris or tissue fluids by sucking (psoroptic mange) or biting (chorioptic mange). Due to mite bites and reaction to saliva, mange is connected with scabs and severe itching (pruritis), which in turn causes damage of the integument due to rubbing and licking. Advanced lesions are described as hairless, scaly and scabby areas and crusts. Preferential locations for mange are the lateral udder in cows and the insides of the hind legs in young stock, as well as the area around the tail and sacrum and at the head and neck (Rosenberger, 1970).

Pediculosis
The ectoparasites causing pediculosis are lice. Distinction is drawn between chewing and biting lice (mallophage) which feed on exfoliated epithelium and skin debris, and sucking lice (anoplura) feeding on blood and tissue fluid. Depending on degree of infestation, hairless patches, skin irritation and chronic dermatitis in association with itching can be observed. Similar to mange, consequential injuries through self-inflicted trauma can be found. Even though pediculosis is supposed to be harmful only when infestation is heavy, hide damage and decreased growth even at lower levels (Rosenberger, 1970) indicate welfare relevance already at this point. Favoured sites of infestation are the neck and the area around the withers of cattle.

In clinical examination, the degree of infestation with ectoparasites is measured by counting individuals on cattle or from a skin scraping (e.g. Rosenberger, 1970; Colwell & Himsl-Rayner, 2002). Coles et al. (2003) assessed damage due to external parasites in the hides after slaughter, but not in a quantitative manner.

Alterations Due to Actions of Pen Mates

Agonistic interactions
While agonistic interactions in dehorned cattle mostly do not lead to overt skin damage, horned cattle can inflict severe injuries. The degree of injuries depends on the form of the horns and temperament of the cows, the suitability of housing conditions, the management and human-animal relationship (Menke et al., 1999). Following Menke (1996), lesions inflicted by horns can be differentiated from lesions due to housing equipment by their vertical direction and are mostly located at the rump of the animals. Groth (1984, 1985) mentions that especially in bulls, horn inflicted lesions can be found at the head.

Stepping on teats or tails
Depending on udder size and shape as well as housing conditions (e.g. Ekesbo, 1966; Blom, 1983; Groth, 1985), injury on teats can occur through stepping of pen mates, or often the cow herself (but udder lesions can also occur after collisions with barbed wire fences on pasture). Resulting teat lesions are described as erosions, contusion and tear-off (partial or total) (Groth, 1984, 1985), or superficial and deep wounds, scab, erosion and loss of teats (Zerzawy, 1989), injuries (Platz et al., 1999; Bareille et al., 2003), lesions (Danuser and Regula, 2001), or tramped teats (Ekesbo, 1966).

Especially in fattening bulls, stepping on the tail-tip can cause cracks, injuries (scab, blood secreted tissue fluid, purulent or necrotic) and swellings (Schrader et al., 2001), or necrosis (Groth, 1984, 1985).

Cross-sucking/inter-sucking
Cross-sucking or inter-sucking, mainly in young animals, but sometimes also in dairy heifers and cows, may cause hairless patches or injuries at the prepuce (Groth, 1984, 1985; Plath et al., 1998), scrotum or teats (Lidfors and Isberg, 2003, review). Due to the fact that ears are among the preferred sites of cross-sucking, alterations may be found also there.

Housing Equipment

In general, most alterations of the integument are caused by repeated collisions or contact with housing structures. They are also called technopathies, and they are most prevalent at leg joints (carpus, fetlock joints, stifle and tarsus), withers, neck, hip and spine/backbone, as well as brisket and shoulders. However, protruding and sharp-edged parts of equipment in the housing system may cause injuries at any part of the body (Groth, 1985).

6.3 RESULTS

6.3.1 Methods of Assessment of Skin Alterations

Using our own data bases and CAB Abstracts for literature searching, we found 37 relevant studies in which alterations of the skin were assessed and that report in sufficient detail about their assessment scheme, with most of them relating to dairy cows. Results of reliability testing are presented in only two studies. Also in two studies, information is given on time requirements of the assessment, although this included additional measures in both cases and did not match the recommended recording scheme in total. No investigation could be found on necessary sample sizes or on the welfare impact of different alterations on cattle. Most of the studies aimed at comparing effects of different housing or management conditions on health of the animals or at determining risk factors for skin alterations.

The different working groups vary considerably with regard to the classification, quantification and welfare evaluation of skin alterations. The easiest method is to only record any alteration as present without any classification (Alban et al., 1996; Hindhede et al., 1996; Mogensen et al., 1997; Bareille et al., 2003). However, often alterations are differentiated in varying detail according to location, type and dimension or severity of alteration. Furthermore, they are either counted, assessed in terms of area affected or allocated a score. Allocating a score includes evaluation of the severity of alterations (less severe alterations are allocated lower scores) and calculation of the degree of damage on the whole animal.

6.3.2 Classification of Alterations

The location of an alteration potentially yields information about its cause. For the reasons stated above, we will disregard location in the following and concentrate on the classification of type and size.

Types of alterations can be hyperkeratosis (=callosity), pale hairless, red hairless, abrasions, open or infected wounds, superficial lesions, open sores, scabs of dried blood, scars, crusty skin, local inflammations, cellulites, abscesses, phlegmone, ulcerations and swellings as described and assessed, with varying grade of differentiation, in different studies (Ekesbo, 1966; Gruner, 1972; Groth, 1984, 1985; Zerzawy, 1989; Rodenburg et al., 1994; Troxler, 1994; Menke, 1996; Friedli et al., 1999; Hindhede et al., 1999; Schaub et al., 1999; Chaplin et al., 2000; Weary and Taszkun, 2000; Wechsler et al., 2000; Danuser and Regula, 2001; Schrader et al., 2001; Vokey et al., 2001; Livesey et al., 2002; Klaas et al., 2003; Whay et al., 2003; Kögler et al., 2004; Zurbrigg et al., 2005; Keil et al., 2006; Schulze Westerath et al., 2007).

In most cases, the size or grade of the alterations was differentiated. For instance, the area affected was recorded or alterations were graded as slight or severe, or even as slight, medium or severe. Often the size of alterations was divided into classes by their diameter, with 2 cm and 4 or 5 cm, respectively, being usual thresholds, but also more detailed differentiations were made (<1 cm, 1 2 cm, 2–4 cm, 4–8 cm, >8 cm).

6.3.3 Quantification of Alterations

Quantifying the amount of alterations is another important aspect in the evaluation of the impact on the animal. As reviewed by Webb and Nilsson (1983), variables of measuring alterations are the frequency and number of alterations, the severity of alterations, the number of types of alterations and the number of sites at which alterations are found. The quantity of alterations can be calculated as number or percentage of animals or sites affected (e.g. Ekesbo; 1966; Gruner, 1972; Zerzawy, 1989; Rodenburg et al., 1994; Chaplin et al., 2000; Weary and Taszkun, 2000; Danuser and Regula, 2001; Livesey et al., 2002; Whay et al., 2003; Kögler et al., 2004; Zurbrigg et al., 2005), or as number of lesions per animal or per location (Troxler, 1994; Friedli et al., 1999; Schaub et al., 1999; Wechsler et al., 2000; Schrader et al., 2001). Occasionally, skin alterations were assessed as area affected (e.g. Weary and Taszkun, 2000; also used by Mowbray et al., 2003, measured by means of the diameter in x- and y-axes).

Some authors scored the alterations, and registered only the most severe score at a given site (e.g. Krebs et al., 2001; Kögler et al., 2004) Another approach is to calculate total alteration scores (Walberg, Blom, both cited in Webb and Nilsson, 1983; Gustafson et al., 1993; Busato et al., 2000; Vokey et al., 2001; Bahrs, 2005; Schulze Westerath et al., 2007).

6.3.4 Welfare Evaluation of Alterations

Some authors used interpretative classifications of the alterations in terms of their welfare impact. For instance, Kögler (2003) judged hairless patches as mild alterations, abrasion of the skin as medium and swellings at the bursa as severe alterations. In a similar way, but only using two categories, Livesey et al. (2002) characterised hair loss as minor abnormality and skin abrasion, subcutaneous swelling and enlargement of the hock joints or associated synovial structures as severe abnormalities. Slightly deviating, Groth (1985) termed not only hair loss, mild abrasion and callosity, but also swelling of the bursa at the leg joints as light alterations. Open infected wounds, phlegmone, bursitis and arthritis were judged as critical and associated with pain. However, none of these evaluations were based on any validity studies.

Only few reference values for a herd welfare assessment on the basis of skin alterations can be found. Bartussek et al. (2000) provide limits regarding the percentages of animals affected by minor to bad degrees of damage in order to evaluate herd welfare as very good to very bad. Platz et al. (1999) within a herd health assessment system allocated lower scores when more than 0 % or 2.5 % of animals showed technopathies. The DLG (2000) defined maximum prevalences of different degrees of tarsal joint alterations in order to evaluate soft lying mats for cubicles as recommendable in terms of animal welfare. However, the scientific basis for the different reference values is not provided or discussed. Moreover, little or no definition of the categories of alteration is given by Bartussek et al. (2000) and Platz et al. (1999).

6.3.5 Reliability

Danuser & Regula (2001) report agreements between two observers of 75 % for swellings at carpus or tarsus (yes – no), 78 % for injuries at the rump (no – one abrasion – more than one abrasion) and 79 % for alterations at the tarsus (no alteration – hairless patch – redness or swelling – open wound or abscess). They investigated 94 animals and used kappa-values for the comparison. Zurbrigg et al. (2005) trained numerous inspectors to assess skin alterations through scoring cards and on farm until an agreement of 80 % with previous scores and with those of other inspectors were reached. Details about methods used (including number of inspectors) and amount of training are not given. However, agreement on recording teat injuries was low (Zurbrigg et al., 2005).

6.3.6 Feasibility

Skin alterations were either evaluated by palpation or by inspection or by both. In most investigations all animals of a herd or batch were examined. Whay et al. (2003) selected 20 % of the cows by examining every fifth animal. In one case where the effects of different lying mats were investigated, focal animals were selected according to a minimum time spent in the housing system and stage of lactation (DLG, 2000).

From only two studies data are available about the necessary time for the assessment of the integument. Zurbrigg et al. (2005) report that half a day was spent for one herd of dairy cows. They assessed the condition of the hock (no alteration – swollen – hairloss – broken skin or scabs) and additionally, other animal-based parameters (presence of an arched back, position of the claws and animal cleanliness) and stall-based measurements were assessed. Krebs et al. (2001) needed no more than three minutes for the examination of a cow including an assessment of cleanliness, status of coat and claws and alterations at the legs (no alteration – loss of hair – swelling or inflammation – presence of open wound or abscess, the most severe was scored) and at other parts of the body (no alteration – loss of hair – superficial wounds – deep wounds or abscesses).

6.4 DISCUSSION

The assessment schemes for skin alterations vary considerably between studies. Methods are specific to the subject of investigation and the animal category (dairy, beef, calves). This leads to differences in body location investigated, in the degree of differentiation concerning type and extent of alterations, and in the sampling methods. No scheme appears to be directly applicable for a welfare assessment scheme covering all aspects of the condition of the integument in cattle. However, broadly used categorisations of alterations are hairless patches, alterations characterised by broken skin in different forms, and swellings, which are judged diversely on their severity. Values often used to distinguish different sizes of skin lesions are 2 cm and 4 or 5 cm. The frequency of animals affected and the mean number of alterations per animal or per location are often calculated. Due to handling problems with bulls or cows in extensive housing conditions, a palpatory assessment of skin alterations is not feasible. With this precondition, solely clearly visible alterations can be assessed. Small alterations, especially at the tail-tip, very likely cannot reliably be recorded.

6.5 CONCLUSIONS

Based on the literature review and our own practical and scientific experience, we propose the following on-farm assessment scheme for skin alterations in cattle:

Animals are visually examined from one side from a maximum distance of 2 m. This allows the inspector to assess bulls without entering the pen and extensively housed cattle without the need for capture. Only alterations with a minimum diameter of 2 cm at the largest extent are recorded, and two categories are distinguished: i. hairless patches (area with hair loss including extensive thinning of the coat as a response to parasites, skin not damaged, hyperkeratosis possible); ii. lesions (damaged skin either in the form of a scab or a wound, dermatitis due to ectoparasites and (partly) missing teats) and overt swellings. Alterations of the two categories are counted per animal and the percentage of animals affected as well as the mean number of alterations per animal calculated for each category. The same sample of animals can be used for the assessment of the integument as for the scoring of cleanliness, body condition and lameness.

The proposed assessment system is easy to learn and simple to apply. We estimate that the assessment takes no longer than 2 minutes per animal depending on the number of alterations, accessibility of the animals and complexity of the farm regarding groups and buildings. Assessors should be trained with the aid of photographs of various affected animals and additionally on farm. Reliability should be tested during training.

7

ASSESSMENT OF ULTIMATE PH AND BRUISING IN CATTLE

R. Westin, A. Velarde, A. Dalmau and B. Algers

7.1 SUMMARY

There are strong relationships between animal welfare, meat quality, ultimate pH and bruising. This makes ultimate pH and assessments of bruise level valid candidates for animal welfare auditing.

Questionnaires have been sent to 10 abattoirs in 3 EU countries (Belgium, Spain and Sweden). The results show that both sampling procedures (point of time for sampling after slaughter, sample site) and acceptable pH range varies (<5.8 up to 6.2) within and between the countries. It was therefore concluded that it is not possible to use the companies own records in a welfare audit.

The Australian Carcass Bruise Scoring System (ACBSS) is concluded to be a good method for scoring of bruising in cattle. The carcase is divided into 7 defined areas (butt, rump and loin, rib, forequarter, back, hip, and pin). In each area, the level of bruising is assessed on the basis of diameter (Slight = 2–8cm, Medium = 8–16cm, Heavy = >16cm) and depth (deep = other than surface muscle involved). Bruises below 2 cm in diameter are not recorded, as well as fire bruises (superficial bleedings in the fat) and bruises caused by shackling. This makes a total of six categories: Slight (S), Slight-deep (Sd), Medium (M), Medium-deep (Md), Heavy (H), Heavy-deep (Hd). Bruise scores are then calculated by multiplying the assessed bruises in each site class by a weighting factor (S=1, Sd=3, M=3, Md=5, H=5, Hd=7) and summing these values.

7.2 INTRODUCTION

Meat pH and bruising are well known parameters of meat quality, closely related to animal welfare and economically valuable (Gregory, 1998; Atkinson, 2000). Carcase bruise marks arise for instance when the live animal strikes against fences or sharp corners causing blood vessels to rupture. Transport and lairage conditions, facility design and animal handling procedures have been shown to influence the level of bruising in cattle (McNally and Warriss, 1996; Weeks et al., 2002).

When an animal is slaughtered, its muscle continues to metabolize energy thru breakdown of glycogen. As glycogen storage decreases, lactic acid instead accumulates causing meat pH level to drop during the early post-mortem period. Normal pH levels drop from pH 6.8 to 5.6 within 24 hours known as the ultimate pH (Atkinson, 2000). Stress in the live animal also leads to depletion of glycogen storage since stress-induced release of adrenalin triggers energy metabolism and the breakdown of glycogen. If the muscle does not contain sufficient levels of glycogen at slaughter, only a small drop in ultimate pH level will occur and the meat is likely to become dark, firm and dry (DFD). Thus, a high pH after 24 hours indicates that the animal has been metabolically stressed before slaughter (Gregory, 1998). Exposing animals to long-term stress conditions i.e. extreme environmental temperatures, being mixed with unfamiliar cohorts, long transportation and lairage times, and prolonged rough handling can contribute to causing DFD in slaughter animals (Atkinson, 2000; Gregory, 1998). There are however several different ultimate pH limits that has been used for assessment of DFD incidence in cattle (Fabianson et al., 1982).

The aim of this study was to examine what pH ranges are considered as acceptable in different EU countries considering meat quality/incidence of DFD in cattle. The objective was also to develop and evaluate a scoring method for assessment of carcase bruising.

7.3 METHODS

Questionnaires have been sent to 10 abattoirs (4 in Belgium, 3 in Spain, 3 in Sweden) with questions regarding accepted meat pH ranges and present measuring procedures for sampling of ultimate pH and bruise carcase bruising. One of the abattoirs was visited during one day when bleedings on carcases were documented on photos. Contact has been made with 3 different Swedish veterinarians working with meat inspection to discuss assessment of bruising. In addition, literature regarding methods for scoring of carcase bruising has been scanned.

7.4 RESULTS AND DISCUSSION

7.4.1 pH

All 10 abattoirs answered the questionnaires. In 9 of the 10 abattoirs pH was regularly measured and controlled after slaughter. Accepted pH ranges and sampling procedures are listed in Table 7.1.

The results in table 1 indicate that there are differences both within and between different European countries regarding acceptable levels of ultimate pH. Three abattoirs accepted a pH of up to 6.2, two accepted pH 6.0, one accepted pH 5.9 and the last three accepted pH 5.8. One abattoir did not use pH as a measure of meat quality at all. Seven abattoirs measured pH in M. Longissimus dorsi.

Meat with a high ultimate pH (DFD meat) is less resistant to microbial spoilage and is less tasty compared to meat with normal pH (Gregory, 1998). Differences in acceptable pH ranges can thus have economically reasons. However, different pH limits ranging from 5.8 to 6.4 have been suggested in the literature (Fabiansson et al., 1984) and still there are different opinions. In a recent survey of 1000 animals a longissimus muscle pH of 5.87 was identified as the approximate cut-off limit between normal and DFD meat (Page et al., 2001). The authors in another study conclude that carcasses that fall into the ultimate pH range of 5.8–6.0 should be regarded as 'DFD' if they are steers or heifers but 'normal' if they are bulls (Wulf et al., 2002). A lot of research investigating the relationship between animal welfare and high pH define DFD as ultimate pH ≥6.0 (Gregory, 1998; Scanga et al., 1998; Apple et al., 2005).

TABLE 7.1 Accepted meat pH range and use pH-sampling procedures for 10 abattoirs.

Abattoir (Country)	Accepted pH range	Muscle used for sampling	Time lag post slaughter	Amount of animals sampled
1. (BEL)	Not measured	–	–	0%
2. (BEL)	5.5<pH<6.2	Gluteus muscle	24h	?
3. (BEL)	<5.8	M. Longissimus dorsi	+/– 48h	50%
4. (BEL)	5.8<pH<6.2	M. Latissimus dorsi	24h	6%
5. (ESP)	4.3 to < 5.9	M. Longissimus dorsi	24h	2%
6. (ESP)	5.5 to < 6.0 Sometimes pH 5.4 is accepted.	M. Longissimus dorsi	24h	20%
7. (ESP)	< 6.0	M. Semi–membranosus	24h	10%
8. (SWE)	< 6.2	M. Longissimus dorsi	12h	100%
9. (SWE)	<5.8 If 5.8–6.2 New sampling is done after 12h.	M. Longissimus dorsi	24h	100%
10. (SWE)	<5.8 If 5.8–6.2 Different parts of the carcase are individually measured.	M. Longissimus dorsi	24h	100%

There is a strong relationship between animal welfare and muscle pH. This makes meat pH a valid candidate for auditing of welfare. However, measurement must be taken after slaughter (24–48 h after slaughter), at the same site and acceptable pH range must be defined. Since all three parameters varies both within and between different EU countries, it is at the moment not possible to use the slaughter companies own records.

7.4.2 Carcass Bruise Scoring

No asked abattoir used any standardized method for scoring of bleedings/bruises on carcase. The carcases are visually inspected and bleedings are trimmed but no documentation is performed unless crime against animal protection legislation is suspected. At Swedish abattoirs, the farmers got compensation for the trimmed meat if the bleedings are pronounced and have arose after the animal left the farm. However, there is no standardized system in use for assess of the age of the bleedings.

In the literature three methods for scoring of bruising has been identified. The most frequently used in research is 'The Australian Carcass Bruise Scoring System' (ACBCC) developed by the Australian Meat and Livestock Corporation and the Queensland Department of Primary Industries (Anderson and Horder, 1979). In this method, bruising is recorded on seven different areas on both sides of the carcase (butt, rump and loin, rib, forequarter, back, hip, and pin). At every area, the level of bruising is assessed on the basis of diameter (<2cm, 2–8cm, 8–16cm, >16cm) and depth (other than surface muscle involved or not). Bruise scores are then calculated by multiplying the assessed bruises in each site class by a weighting factor (1,3,5 or 7) and summing these values. Several different possibilities for comparison of bruise levels are than available i.e. mean bruise score per carcase, distribution of bruising and proportion of bruised animals (Wythes et al., 1985; Weeks at al., 2002). The second identified method, devised by McNally and Warriss (1996), is very similar to ABCSS but assessment is based on area instead of diameter (2–400cm^2, 400–800cm^2, 800–1200cm^2…) and the depth of the bruise is divided into 3 categories (<1cm, 1–2cm, >2cm) instead of two. Bruise score in each site class are then calculated from the bruised area divided with 400 and multiplied with a weighting factor (1, 2 or 3). Summation of these values will than give the average bruise score for the whole carcase. The third described method is used in National Beef Quality Audits in North America. Five or seven different sites are studied (round, loin, rib, chuck, brisket, flank and plate). The severity of a bruise is subjectively scored into 3 or 4 categories (minor, medium, major, extreme/critical) reflecting the weight of trim needed to remove the bruise (Roeber et al., 2001; Bolemanet et al., 1998; McKenna et al., 2002). The scoring result is used for analysis of severity of bruises within site but no average score for the whole carcase is calculated. Instead the number of bruises on carcase is used for comparison between carcasses.

All of these methods are quite similar. The carcase is divided into several different areas where bruises are individually scored. Scoring is based on spread and depth of the bruise, or the weight of trim needed to remove the bruise. In ACBSS and the method devised by McNally and Warriss (1996) a total carcass score is than calculated by adding the subscores from each recorded site, making it possible to calculate a total score of the whole carcase and an average mean score within all monitored carcasses. One of the advantages with dividing the carcases into different defined sites is that different problem areas related to animal welfare can be identified. This kind of information can be very valuable and is useful in a company's bruise preventive work (Haeger, personal message, 2006). According to the original proposal, observers were supposed to assess carcases bruise level on a scale from 0–3. If assessment of bruise level is done without considering the location of the bruises, the possibility of identifying specific problem areas related to animal welfare is lost. Such a scoring system is therefore not preferable (Haeger, personal message).

The Australian carcase bruise scoring system (ACBSS) is well documented and has been used in several abattoir surveys in both Australia and Europe (Jarvis et al., 1995; Wythes et al., 1985; Weeks et al., 2002). We therefore conclude that this scoring method is well suited for scoring of bruise level within the Welfare Quality project and by using this method, previous research results can be used as reference. Since the method is previously tested and evaluated (Anderson and Horder, 1979) no inter-observer reliability tests has been performed as proposed in the original proposal.

7.5 CONCLUSION

Both ultimate pH and scoring of bruising in cattle are valid candidates related to animal welfare. However, sampling of pH must be done at the same time after slaughter, at the same site and a standardized acceptable pH range must be defined. Since all three parameters varies both within and between different EU countries, it is not possible to use the slaughter companies own records. 'The Australian Carcass Bruise Scoring System' is concluded to be a good method for assessment of bruising.

8

RELIABILITY OF MEASURES OF INJURIOUS AND ABNORMAL BEHAVIOUR IN DAIRY AND BEEF CATTLE

N. Brörkens, S. Laister, S. Lolli, D. Zucca, C. Winckler, M. Minero, E. Canali and U. Knierim

8.1 SUMMARY

Injurious behaviours such as slipping, falling or stepping on other group mates may lead to wounds and skeletal or muscle problems. Reduced welfare in cattle is also assumed with the occurrence of stereotypies such as tongue rolling or redirected behaviours such as intersucking or chewing of objects in cattle.

It was the aim of this study to investigate injurious and abnormal behaviours as on-farm measures of welfare in dairy cows and fattening bulls with regard to feasibility, inter-observer reliability and short- to long-term intra-farm variability (consistency). For this purpose, continuous behaviour sampling was carried out on three days for 4 h each on 43 dairy (20 cubicle houses, 11 deep litter systems, 12 tie stalls) and 20 beef farms (10 deep litter, 10 fully slatted floors) in Austria, Germany and Italy. In beef farms, three weight classes were defined (200–350 kg, 350–500 kg, >550 kg). Farm visits took place at approximately 60 and 180 days (tie stalls: 60 and 120 days) after the first visit. Inter-observer reliability was tested during direct observations (6 dairy, 2 beef farms) and using video clips (n=55).

In this study abnormal behaviours occurred only rarely during the 4 hours of on-farm observations (median events/animal*hour = 0.00) in both dairy and beef cattle farms. Abnormal behaviours therefore have not been considered in further statistical analysis and can not be recommended as reliable parameters for on-farm welfare assessment in cattle.

Only in beef bulls, the injurious behaviours slipping (median: 0.13–0.24 events/animal*hour), mounting (median: 0.19–0.28 events/animal*hour) and total injurious behaviour (median 0.44–0.60 events/animal*hour) occurred sufficiently frequent

for reliable recordings. Due to low incidences inter-observer-reliability analysis was only possible for slipping and only from video observations, but revealed unsatisfactory agreement between observers (Kendall's W= 0.64).

Regarding consistency over time, slipping reached acceptable levels only when medium and finishing bulls were pooled (W=0.72). Furthermore, mounting and the total of all injurious behaviours were not found to be consistent over time regardless of the weight of the animals (W<0.70).

In conclusion, none of the behaviours studied is suggested for inclusion in an on-farm welfare assessment protocol for dairy or beef cattle.

8.2 INTRODUCTION

Injurious behaviours such as slipping, falling or stepping on other group mates may lead to wounds and skeletal or muscle problems (van der Tol et al., 2005). For example, trampling by pen mates is regarded a major cause of tail tip alterations in beef bulls (Schrader et al., 2001).

With regard to abnormal behaviours in cattle, stereotypies such as tongue rolling or redirected behaviours such as inter-sucking or chewing of objects have been described as potential indicators for impaired welfare in calves (Scientific Veterinary Committee 1995; Rushen and de Passillé, 1995), heifers (Redbo, 1992; Redbo et al., 1996; Redbo and Nordblad, 1997), dairy cows and fattening animals (Sambraus and Gotthardt, 1985; Graf, 1994).

The objective of this study was to investigate injurious and abnormal behaviours as on-farm measures of welfare in dairy cows and fattening bulls with regard to feasibility, inter-observer reliability and short- to long-term intra-farm variability (consistency).

8.3 METHODS

8.3.1 Inter-observer Reliability Testing (IORT)

IORT of injurious and abnormal indicators was based on (1) direct observations on six dairy and two beef farms and (2) recordings from video clips. For the investigation of inter-observer reliability no distinction was made between dairy cows and fattening bulls as behaviour patterns are similar in both cattle categories.

On-farm IORT

On-farm IORT was carried out three times at different stages of the project with different numbers of observers being present at each date (Table 8.1). The first testing took place after two days of initial training before on-farm data collection started. The training included discussions, video and on-farm training. The second IORT was about 50 days after observers A and B had started on-farm data collection; two additional observers (C and D) participated, who had received about half a day of training. The final on-farm IORT took place after data collection had been finished.

Observers were always located near to each other in the barn allowing a free view on the area observed.

Spearman rank correlation (r_s) was used to test agreement within pairs of observers. Additionally Kendall's coefficient of concordance (W) was calculated for agreement between 3 and 4 observers, respectively.

Table 8.1 Overview of on-farm IORT.

Meeting	mm/yyyy	Number of farms visited	Sample size (pens, segments)			
			A	B	C	D
1 Germany	07/2005	4	35	35	–	–
2 Austria	09/2005	2	10	10	7	10
3 Italy	06/2006	2	15	15	15	–
	Total	8	60	60	22	10
	Dairy	6	50	50	22	10
	Beef[1]	2	10	10	–	–

Notes: [1] 07/2005; A, B, C, D = observers.

IORT Using Videos

After completion of the on-farm data collection and after the final on-farm IORT in June 2006, three trained observers (A, B & C) separately watched 55 video sequences of about 2 to 14 minutes (07:05 hours in total). The video clips had been recorded on different beef and dairy farms and contained representative situations of the behaviours that were in the scope of the study. The video observations followed the same rules as provided in the instructions for data collection.

Again Spearman rank correlations and Kendall's coefficient of concordance (W) were used.

8.3.2 Investigation of Intra-farm Consistency

In total, 43 dairy farms (20 cubicle, 11 deep litter, 12 tie stall systems; herd size 12 – 150 cows) and 20 beef fattening farms (10 deep litter, 10 fully slatted floor systems; animals per farm: 30-220, 5-27 bulls per pen) in Austria, Germany and Italy (only dairy) were included in the study (Table 8.2). The dairy cows belonged to different breeds with Holstein Friesian, Simmental-Fleckvieh and Brown Swiss being the most prevalent breeds. The fattening bulls were Simmental-Fleckvieh (S-FV), Limousin and S-FVxLimousin crosses. Other breeds such as Belgian Blue, Brown Swiss, Holstein Friesian, Tyrolean Grey or Charolais were also kept in small numbers.

Observations of injurious and abnormal behaviours as defined in Table 8.3 were carried out on three days at intervals of approximately 60 and 180 days after the first visit (beef cattle farms, dairy loose housing systems). In dairy farms with tie stalls, the third visit took place 120 days after the first visit (Figure 8.1).

Table 8.2 Overview of number and type of farms visited in each country.

	Austria	Germany	Italy	Total
DAIRY				
Cubicles	8	8	4	20
Deep litter	3	4	4	11
Tie stalls	6	6	–	12
Total	17	18	8	43
BEEF				
Fully slatted	5	5	–	10
Deep litter	5	5	–	10
Total	10	10	–	20

TABLE 8.3 List of behaviours observed and their definitions.

Behaviour	Definition
Slipping (SP)	One or more claws accidentally slide out of place or glide off edges and/or the animal is losing balance.
Falling (FA)	An animal accidentally looses balance and its body quickly moves towards the ground and touches it with udder, sternum, carpal joint, knee or with the whole flank or abdomen. If falling is caused by previous slipping then slipping is counted separately.
Stepping on… (only beef bulls)	… tail (ST): A standing or walking bull puts its foot on the tail of a lying group mate and puts weight on it. Each step is counted. Only tested in fully slatted housing systems.
Stepping on… (only beef bulls)	… other parts of the body than tail (SB): A standing or walking bull puts its foot on any part of the body other than tail of a lying group mate and puts weight on it. Each step is counted.
Mounting (MO) (only beef bulls)	A bull lifts itself up on its hind legs and jumps with its forelegs onto another group mate either from behind, the side or front. The receiver may be standing or lying.
Injurious total (INJ)	Dairy cows: INJ = SP + FA Beef bulls: INJ = SP + FA + ST + SB + MO
Tongue rolling (TR)	The animal is repeatedly twisting, twirling or swinging the tongue in a stereotypic way inside or outside the open mouth, or stretches out the tongue for longer than 10 seconds; sometimes the neck and head is stretched upwards. A new bout is recorded if the behaviour is interrupted for more than 10 s.
Chewing on equipment (CW)	The animal takes any equipment in its mouth and chews on it for more than 10 seconds. A new bout is recorded if chewing is interrupted for more than 10 seconds.
Intersucking (IS)	The actor gets hold of teat, udder, ear, tail, prepuce or skin fold of a group mate with its mouth and pulls at it with the muscles of its cheeks and tongue as if it wants to get milk out of it.
Sitting (SI)	The animal is sitting dog-like on its hind quarter with front legs fully stretched for longer than 10 seconds. A new bout is recorded when the animal has either adopted another lying position or risen in between.
Abnormal total (ABN)	ABN = TR + CW + IS + SI

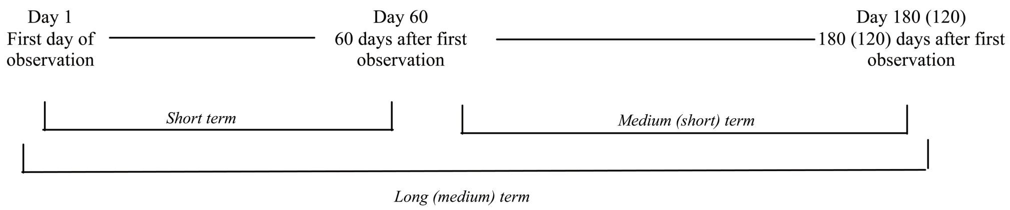

FIGURE 8.1 Schedule for on-farm observations.

Data Recording on Dairy Farms

On each farm the lactating dairy cows were observed excluding separate groups of dry or periparturient cows, as well as cows in hospital pens. Behaviour performed by or with animals in heat was not recorded.

Behaviour was recorded using continuous behaviour sampling and observations lasted for 4 h after morning feeding (or after feed had been pushed up). The observer was positioned on the feed bunk on an elevated observation chair.

In herds larger than 25 cows, the observations were carried out in segments of the barn which were expected to contain on average not more than 25 cows per segment. These segments covered all accessible areas (lying areas, feeding places, concentrate feeders, outdoor loafing areas etc.). The duration of continuous observations within each segment was adjusted to the number of segments so that each segment was observed at least once per two hours (minimum observation period 10 min) or at least twice within four hours respectively.

The number of animals which were feeding, standing/walking and lying within the segment was recorded at the beginning and at the end of each observation period. Data were then analysed as the mean number of events per animal and hour, taking the absolute frequency of behaviours, the duration of observations per pen/segment and the average number of animals in the pen/segment during the observation into account.

From all values obtained on segment level the mean incidence at herd level was calculated (occurrence of behaviour/animal*hour). Spearman rank correlations and Kendall's coefficient of concordance were used to test consistency between visits (1–60–180/120).

Data Recording on Beef Farms

Three weight classes were defined in line with the literature and common farming practice:

- initial fattening period (I): 200–350 kg;
- medium fattening period (M): 350–550 kg;
- finishing fattening period (F) ≥550 kg.

Pens holding less than three animals were excluded from observations.

Behaviour was recorded in the same way as on the dairy farms.

All weight classes present were observed for equal periods within each observation hour and each pen was observed at least twice during the 4 hour period. Pens with more than 25 bulls were divided into 2 or more segments (see dairy cattle).

Also data processing within each weight class followed the same rules as for dairy cattle.

8.3.3 Decision on Measures

We followed a stepwise approach in deciding on the usefulness of the behavioural measures for on-farm welfare assessment protocols:

1. Median on-farm incidence had to exceed 0.10 events/animal*hour to allow for reliable and feasible recording as well as differentiation between farms;
2. In terms of inter-observer reliability (IOR), correlation between observers had to be rs or Kendall's W≥0.70.
3. In terms of intra-farm consistency, again for the remaining measures correlation coefficients between results from different observation days had to be Kendall's W≥0.70.

8.4 RESULTS

8.4.1 Incidences of Agonistic Social Behaviours

In dairy cows, all injurious or abnormal behaviours occurred at median incidences of less than 0.10 events/animal*hour (Table 8.4). Median incidences of both total injurious and total abnormal behaviours also did not exceed the threshold of 0.10 events/animal*hour.

In beef bulls, injurious behaviours occurred more frequently and the median level of total injurious behaviour ranged from 0.44 to 0.60 events/animal*hour (Table 8.4). Mounting was the injurious behaviour most often observed and was most frequent in medium fattening bulls (0.28 events/animal*hour). Incidences of slipping were highest in finishing (0.24 events/animal*hour) and lowest in initial fattening bulls (0.13 events/animal*hour). Median incidences of other injurious behaviours (falling, stepping on tail, stepping on bull) were always below 0.10. Similar to dairy cows, the median frequency of abnormal behaviours in beef bulls never exceeded 0.00 events/animal*hour.

Box 8.1 Measures of injurious and abnormal behaviour that are feasible for on-farm welfare assessment.

Dairy cows	*Beef bulls*
Injurious: none	Injurious:
Abnormal: none	Slipping
	Mounting
	Total injurious behaviour
	Abnormal: none

TABLE 8.4 Descriptive measures (median, mean, minimum (Min) and maximum (Max) values, standard deviation of the mean (SD) and variance (Var)) of incidences of injurious and abnormal behaviours in dairy cows and beef bulls (events/animal*hour); for beef cattle overall data for the three farm visits are presented.

Behaviour	Animals	Day Weight	Median	Mean	Min	Max	SD	Var	N
	DAIRY Loose housing	1	.02	.09	.00	.70	.17	.03	31
		60	.03	.08	.00	.48	.12	.01	31
		180	.03	.05	.00	.41	.09	.01	31
		overall	.03	.07	.00	.70	.13	.02	93
	DAIRY Tie stalls	1	.01	.03	.00	.14	.05	.00	12
		60	.04	.04	.00	.15	.05	.00	12
		120	.02	.03	.00	.11	.04	.00	12
		overall	.03	.03	.00	.15	.04	.00	36
	BEEF	I	.13*	.26	.00	1.41	.33	.11	56
		M	.18*	.46	.00	2.34	.56	.31	57
		F	.24*	.39	.00	3.45	.62	.38	57
Falling (FA)	DAIRY Loose housing	1	.00	.00	.00	.06	.01	.00	31
		60	.00	.00	.00	.02	.00	.00	31
		180	.00	.00	.00	.05	.01	.00	31
		overall	.00	.00	.00	.06	.01	.00	93
	DAIRY Tie stalls	1	.00	.00	.00	.00	.00	.00	12
		60	.00	.00	.00	.00	.00	.00	12
		120	.00	.00	.00	.00	.00	.00	12
		overall	.00	.00	.00	.00	.00	.00	36
	BEEF	I	.00	.01	.00	.15	.03	.00	56
		M	.00	.01	.00	.18	.03	.00	57
		F	.00	.01	.00	.19	.03	.00	57
Stepping on tail (ST)	BEEF	I	.00	.01	.00	.19	.04	.00	56
		M	.00	.00	.00	.05	.01	.00	57
		F	.00	.00	.00	.15	.02	.00	57
Stepping on other parts of the body than tail (SB)	BEEF	I	.00	.04	.00	.23	.07	.01	56
		M	.00	.03	.00	.23	.07	.00	57
		F	.00	.05	.00	.23	.08	.01	57
Mounting (MO)	BEEF	I	.22*	.37	.00	2.46	.51	.26	56
		M	.28*	.37	.00	1.65	.36	.13	57
		F	.19*	.33	.00	2.54	.45	.21	57
Injurious total (INJ)	DAIRY Loose housing	1	.04	.09	.00	.70	.17	.03	31
		60	.03	.09	.00	.48	.12	.01	31
		180	.03	.05	.00	.41	.09	.01	31
		overall	.03	.08	.00	.70	.13	.02	93
	DAIRY Tie stalls	1	.01	.03	.00	.14	.05	.00	12
		60	.04	.04	.00	.15	.05	.00	12
		120	.02	.03	.00	.11	.04	.00	12
		overall	.03	.03	.00	.15	.04	.00	36
	BEEF	I	.51*	.66	.00	2.46	.55	.30	56
		M	.60*	.85	.00	3.45	.71	.50	57
		F	.44*	.74	.00	4.20	.85	.72	57

Notes: * exceed the threshold of 0.1 events/animal*hour

TABLE 8.4 CONT. Descriptive measures (median, mean, minimum (Min) and maximum (Max) values, standard deviation of the mean (SD) and variance (Var)) of incidences of injurious and abnormal behaviours in dairy cows and beef bulls (events/animal*hour); for beef cattle overall data for the three farm visits are presented.

Behaviour	Animals	Day Weight	Median	Mean	Min	Max	SD	Var	N
Tongue rolling (TR)	DAIRY Loose housing	1	.00	.01	.00	.27	.05	.00	31
		60	.00	.01	.00	.08	.02	.00	31
		180	.00	.01	.00	.08	.02	.00	31
		overall	.00	.01	.00	.27	.03	.00	93
	DAIRY Tie stalls	1	.00	.04	.00	.33	.10	.01	12
		60	.00	.03	.00	.35	.10	.01	12
		120	.00	.03	.00	.21	.06	.00	12
		overall	.00	.03	.00	.35	.09	.01	36
	BEEF	I	.00	.24	.00	2.88	.49	.24	56
		M	.00	.11	.00	1.28	.25	.06	57
		F	.00	.10	.00	.63	.18	.03	57
Chewing on equipment (CW)	DAIRY Loose housing	1	.00	.00	.00	.04	.01	.00	31
		60	.00	.00	.00	.03	.01	.00	31
		180	.00	.00	.00	.00	.00	.00	31
		overall	.00	.00	.00	.04	.01	.00	93
	DAIRY Tie stalls	1	.00	.00	.00	.04	.01	.00	12
		60	.00	.00	.00	.00	.00	.00	12
		120	.00	.00	.00	.00	.00	.00	12
		overall	.00	.00	.00	.04	.01	.00	36
	BEEF	I	.00	.02	.00	.25	.06	.00	56
		M	.00	.02	.00	.29	.06	.00	57
		F	.00	.02	.00	.30	.06	.00	57
Intersucking (IS)	DAIRY Loose housing	1	.00	.00	.00	.00	.00	.00	31
		60	.00	.00	.00	.00	.00	.00	31
		180	.00	.00	.00	.00	.00	.00	31
		overall	.00	.00	.00	.00	.00	.00	93
	DAIRY Tie stalls	1	.00	.00	.00	.00	.00	.00	12
		60	.00	.00	.00	.00	.00	.00	12
		120	.00	.00	.00	.00	.00	.00	12
		overall	.00	.00	.00	.00	.00	.00	36
	BEEF	I	.00	.02	.00	.30	.05	.00	56
		M	.00	.00	.00	.15	.02	.00	57
		F	.00	.00	.00	.10	.01	.00	57
Abnormal total (ABN)	DAIRY Loose housing	1	.00	.01	.00	.27	.05	.00	31
		60	.00	.03	.00	.21	.04	.00	31
		180	.00	.02	.00	.14	.03	.00	31
		overall	.00	.02	.00	.27	.04	.00	93
	DAIRY Tie stalls	1	.00	.04	.00	.33	.10	.01	12
		60	.00	.04	.00	.35	.10	.01	12
		120	.00	.03	.00	.21	.06	.00	12
		overall	.00	.03	.00	.35	.09	.01	36
	BEEF	I	.06	.28	.00	3.25	.53	.28	56
		M	.00	.14	.00	1.28	.26	.07	57
		F	.00	.12	.00	.63	.19	.04	57

In conclusion, measures of injurious and abnormal behaviour that are feasible for on-farm welfare assessment are presented in Box 8.1.

8.4.2 Inter-observer Reliability

On-farm Observations

Due to low or zero occurrences of behaviours, measures of inter-observer agreement could not be calculated for any of the behaviours investigated.

Video Observations

On the 55 video clips, 16, 13 and 15 slipping events were recorded by observers A, B and C, respectively. Rank correlations between single observers as well as Kendall's W for overall agreement did not reach acceptable levels (r_s and W<0.7; Table 8.5). Again, IOR for mounting could not tested using video recordings due to low incidence of the behaviour.

8.4.3 Consistency over Different Periods of Time (Intra-farm Variability)

Due to the very low on-farm incidences of the behaviours falling, stepping on tail, stepping on bull as well as of abnormal behaviours (Table 8.4), results on consistency are only presented for slipping, mounting and for total injurious behaviours in fattening bulls (Table 8.6). Spearman rank correlations between single farm visits for slipping were generally low (r_s=0.08-0.59). The consistency regarding levels of slipping behaviour was higher in

Table 8.5 Spearman rank correlation coefficients (r_s) and Kendall's coefficient of concordance (W) for inter-observer agreement based on video clips (n=55).

Behaviour		Observer pairs			Kendall's W
		A_B	A_C	B_C	
Slipping (SP)	r_s	.57	.34	.46	.64
	p	.000	.011	.000	.000

Box 8.2 Measures of injurious and abnormal behaviour that revealed sufficient inter-observer reliability.

Dairy cows	*Beef bulls*
None	Injurious: none (only slipping tested)

TABLE 8.6 Spearman rank correlation coefficients (r_s) for short- (1_60, 60_120), mid- (60_180, 1_120) and long-term (1_180) intra-farm consistency and Kendall's coefficient of concordance (W) for slipping, mounting and total injurious behaviour in fattening bulls of different weight classes (I=200–350 kg, M=350–550 kg, F=>550 kg).

Behaviour	Weight class		1_60	60_180	1_180	Kendall's W
Slipping (SP)	I	r_s	.35	.40	.08	.52
		p	.150	.096	.737	.031
	M	r_s	.50	.51	.59	.68
		p	.035	.009	.031	.000
	F	r_s	.50	.56	.52	.67
		p	.035	.017	.023	.000
	M & F[1]	r_s	.47	.70	.62	*.72*
		p	.051	.001	.005	.000
Mounting (MO)	I	r_s	.31	.41	.10	.54
		p	.210	.087	.693	.019
	M	r_s	.03	.13	.29	.43
		p	.895	.615	.235	.149
	F	r_s	.06	–.23	.19	.32
		p	.812	.348	.443	.534
	M & F[1]	r_s	.13	.06	.29	.44
		p	.598	.801	.226	.128
Injurious total (INJ)	I	r_s	.36	.32	.04	.51
		p	.144	.192	.884	.036
	M	r_s	.28	.25	.31	.51
		p	.263	.317	.199	.034
	F	r_s	.42	.39	.52	.62
		p	.085	.115	.021	.002
	M & F[1]	r_s	.41	.42	.56	.64
		p	.088	.086	.012	.001

Notes: [1] M & F = all bulls > 350 kg (merging data from pens with medium and finishing weight class bulls); figure in italics exceeds threshold of W=0.70.

medium and finishing bulls than in light bulls, but reached acceptable levels only when medium weight and finishing bulls were pooled (M&F, W=0.72).

Consistency regarding levels of mounting was even lower than for slipping and overall consistency (Kendalls' W) was always below 0.70. Lumping all injurious behaviours together (injurious total) did not improve consistency levels to acceptable levels in the different weight classes or when medium weight and finishing bulls were pooled.

In conclusion, no measures of injurious and abnormal behaviour revealed sufficient consistency over time in both dairy and beef cattle.

8.5 DISCUSSION

In the dairy farms studied, incidences of injurious and abnormal behaviours were very low and also the totals of the single measures (injurious total, abnormal total) did not reach the set minimum incidence level for reliable recordings of 0.1 events/animal*hour. No behaviours can therefore be recommended for on-farm welfare assessment in dairy cattle.

In beef bulls most parameters also occurred rarely during the 4 hours of on-farm observations. However, slipping (SP) and mounting (MO) were more frequently recorded (median: SP: 0.13–0.24 events/animal*hour; MO: 0.19–0.28 events/animal*hour). The lowest frequency of mounting was found in finishing bulls. This was probably caused by horizontal bars above the finishing bull pens in some farms which inhibit jumping onto each other.

Although median frequency of falling, stepping on tail or stepping on other parts of the body than tail was 0.00, reasonable peak values were obtained in single farms (maximum values between 0.05 and 0.23 events/animal*hour). Lumping all injurious behaviours together led to median incidences of total injurious behaviour of 0.44–0.60 events/animal*hour in the different weight classes thus allowing reliable recordings.

Due to low incidences of behaviours during both on-farm observations as well as when behaviours were recorded from video clips, investigation of inter-observer-reliability could only be performed with regard to slipping. The unsatisfactory agreement (Kendall's W= 0.64). might be due to difficulties in recognizing the rather quick movements of the legs in a less visible place (poorly lit floor of the pens). Furthermore, the term '[...] and/or the animal is losing balance' might have led to different assessments during the IOR sessions.

Although inter-observer agreement was poor (slipping) or missing (mounting), intra-farm consistencies were calculated assuming that conspicuous behaviours such as mounting should be reliably observed and inter-observer reliability for slipping is likely to be improved by refining the definition or proper training.

However, overall consistency of levels of injurious and abnormal behaviours was generally below the threshold of W=0.7 and was not markedly improved when behaviour classes were lumped together or weight classes in beef bulls were merged.

8.6 CONCLUSIONS

Due to low levels of occurrence and unsatisfactory consistency over time no measures of injurious and abnormal behaviour are recommended for inclusion in on-farm welfare assessment protocols for both dairy and beef cattle.

9

ASSESSMENT OF FEAR AND INJURIOUS BEHAVIOURS AT SLAUGHTER

R. Westin, A. Velarde, A. Dalmau and B. Algers

9.1 SUMMARY

Pre slaughter handling can be very frightening and stressful to animals. A scoring method for assessment of behaviours related to fear and injury has been developed and evaluated. The presented method is a promising scoring method for assessment of behaviours during driving of cattle into the stun box. However, final conclusions about the feasibility of this method for scoring during unloading of cattle can not be made until the effect of an increased training on inter-observer agreement is evaluated.

During sampling, the following behaviours are recorded.

- Run (Rn) – The animal runs, by itself or as a reaction to the handler driving it forward.
- Move backwards/Turn around (Mb) – The animal move backwards, turn around or tries to turn around, by itself or as a reaction to the handler trying to drive it forward.
- Freeze (Fr) – The alley is free but the animal refuses to continue in spite of the handler physically trying to drive it forward. An animal that stops but continues to walk when the handler drives it forwards is not freezing.
- No locomotion behaviour occurred (No) – The animal does not 'run', 'move backwards' or 'freeze'.
- Slip/Fall – The animal looses its foothold.
- Vocalize – The animal vocalizes in any way.

At unloading, sampling is done at group level. The group of animals unloaded together is monitored from walking of the transport vehicle until entering the stable building or gathering pen. If one animal in the group displays any of the defined behaviours, this is recorded in the audit protocol (Appendix 1). When monitoring cattle at driving into stun box, sampling is done at individual level when an animal is driven into the stun box.

Monitoring is performed until the animal is stunned. As soon as the first observed animal is stunned, observation of the next animal in front of the line starts.

From the scoring results, 'the percentage of unloadings where animals display behaviours related to fear or injury' and 'the percentage of animals driven into stun box that display behaviours related to fear or injury' can be calculated.

9.2 INTRODUCTION

Facility design and animal handling practice are well known to influence fear level and the behaviour of cattle at slaughter plants (Gregory, 1998; Weeks et al., 2002; Grandin, 2005). Distractions which impede animal movement and poor maintenance of facilities such as worn, slick floors are identified as two common causes of animal welfare problems in slaughter plants (Grandin, 1996). Animals will often stop and even refuse to move through a handling system if there are distractions such as sparkling reflections, air blowing towards the animals, high pitched noise and movement up ahead of the approaching animals. These distractions can ruin the performance of a well designed facility (Grandin, 1996; Gregory 1998). Vocalization has also been shown to be strongly associated with stressful events at pre-slaughter handling, such as electric prodding, slipping, missed captive bolt stun and excessive pressure exerted on the animal's body by a restraining device (Grandin, 1998). Several restaurant companies have been auditing animal welfare in slaughter plants since 1999 which has resulted in great improvements in pre-slaughter handling of cattle (Grandin, 2005). There are also examples of beef producing companies owning and running abattoirs that have implemented their own welfare audit programs (Jonsson, personal message, 2006). The aim of this study was to develop and evaluate a method for assessment of fear and injurious behaviours at unloading and driving of cattle at abattoirs, possible to conduct during commercial conditions.

9.3 METHODS

Visits have been made in two Swedish and one Spanish abattoir. During each visit cattle were observed and filmed during unloading and driving into the stunning box. After this, discussions between the collaborators were held and injurious behaviours (slip/fall) and behaviours indicating fear (run, move backwards, freeze, vocalize) possible to monitor during commercial conditions were defined. (For definitions, see Chapter 3.1) From collected material, one film exemplifying defined signs of fear and injurious behaviours and another showing continues sequences of unloading and driving of animals into the stun box were made.

Six observers were trained by reading written instructions of how to monitor the defined behaviour at unloading and driving of cattle and by looking at the film with examples of the behaviours. After training, monitoring was done individually by each observer for two days in a row by looking at 18 film sequences of unloading of animals from transport vehicles (in total 113 animals) and 12 sequences of driving animals into the stun box (in total 12 animals). When monitoring of unloading, animals unloaded together in a group were regarded as one observation unit and if one or more animals in the group displayed any of the defined behaviours (run, move backwards, freeze, slip/fall, vocalize) this was recorded in the audit protocol (Appendix 1). In addition the number of cattle in the group monitored was counted. If an animal was prodded or hit any interior structure comments were made. When monitoring driving of cattle into the stun box only the animal in front of the line was observed.

Fleiss' kappa coefficients (Fleiss, 1971) for inter-observer reliability during monitoring of unloading respectively monitoring of driving into stun box were calculated separately for each of the studied behaviours (run, move backwards, freeze, slip/fall and vocalize) by using the 'Attribute Agreement Analysis' function in Minitab (MINITAB® Release 14.1, Minitab Inc, 2003). During monitoring of unloading, no animal vocalized and during monitoring of driving into stun box no animal did run or slip/fall. No statistic analysis was therefore possible to perform for these behaviours.

9.4 RESULTS

In total 30 observations including 125 animals were made by all six observers. During monitoring of unloading, group size differed from 2 to 13 animals. Kappa values and the number of inter-observer agreements day 1 and 2 are shown in Table 9.1. For monitoring of unloading, the lowest value of kappa was obtained for observation of move backwards (0.44) and the highest kappa value was obtained for observation of slip/fall (0.64). For monitoring of driving into stun box the lowest kappa value was obtained for freeze (0.70) and the highest for move backwards (0.95).

One animal was forced to walk forward by use of an electrified prodder, which all observers commented.

TABLE 9.1 Number of inter-rater agreements within 6 observers and corresponding kappa values for studied behaviours.

		Unloading		Driving into stun box	
		No. agreements[1]	Fleiss' kappa	No. agreements[2]	Fleiss' kappa
Run	Day 1	10 (55.6%)	.56	–[3]	–
	Day 2	10 (55.6%)	.60		
	(*mean*)		(*.58*)		
Move backwards	Day 1	9 (5.0%)	.42	10 (83.3%)	.89
	Day 2	9 (5.0%)	.46	12 (100%)	1.00
	(*mean*)		(*.44*)		(*.95*)
Freeze	Day 1	10 (55.6%)	.47	10 (83.3%)	.76
	Day 2	14 (77.8%)	.44	8 (66.7%)	.64
	(*mean*)		(*.46*)		(*.70*)
Slip/Fall	Day 1	11 (61.1%)	.70	–[3]	–
	Day 2	8 (44.4%)	.58		
	(*mean*)		(*.64*)		
Vocalize	Day 1	–[3]	–	11 (91.7%)	.84
	Day 2			12 (100%)	1.00
	(*mean*)				(*.92*)

Notes: [1] number of exact agreements within all observers in 30 observations including 113 animals; [2] number of exact agreements within all observers in 12 observations including 12 animals; [3] no animal displayed the actual behaviour.

9.5 DISCUSSION

Running, moving backwards and freezing are all behaviours associated with locomotion displayed when an animal tries to escape from or avoid a frightening situation (Gregory, 1998). Cattle are herd animals and the behaviour of one animal will often influence other animals in the herd. If a group of animals are moving and one starts to run, the others are thus likely to follow. At most abattoirs, animals are unloaded in small groups. In our opinion, sampling of these behaviours at group level is therefore more informative and more related to animal welfare than sampling on individual basis. Animals vocalizing or slipping/falling are not believed to influence other animals in the same extent. This means that they can be counted to assess the number of injurious and stressful event on individual level as well. To make monitoring of all behaviours equal at unloading, we did however not choose to record these behaviours at individual level. But, it is possible that a well experienced observer would be able to record vocalize and slip/fall at individual level at the same time as running, moving backwards and freezing is recorded at group level.

The value of kappa evaluates the strength of agreement between observers. It has a maximum value of 1.00 when agreement is perfect, a value of zero indicates no agreement better than chance (Fleiss, 1971). The strength of agreement for a value of kappa between 0 and 1.00 can be interpret as poor' if less than 0.20, 'fair' if 0.21–0.40, 'moderate' if 0.41–0.60, 'good' if 0.61–0.80 and 'very good' if 0.81–1.00 (Altman, 1991). According to this grading system the received kappa values in this study indicates a good to very good inter-observer agreement when monitoring driving of animals into the stun box but only an on average moderate inter-observer agreement when monitoring unloading. At monitoring of driving into the stun box one single animal is observed, while during monitoring of unloading several animals are studied at the same time. Group size during this study varied from 2 up to 13 animals. If the group size is large we can expect simultaneous assessment of several behaviours to be complicated resulting in lower inter-rater agreement. However better agreement than the one obtained is preferable in order to conclude that measuring should be done in this way. During monitoring of unloading the observers where instructed to count all animals in the group observed. This can draw attention from assessment of displayed behaviours. During live observations at abattoirs information about how many animals are unloaded from a truck can instead be obtained from the driver of each unloaded transport vehicle which would possibly facilitate sampling. Another possible cause of low inter-observer agreement is insufficient training. No training was conducted where observers and instructor watched video sequences together and discussed interpretation of borderline cases. We believe that a more thorough training can increase inter-observer agreement. This should be evaluated before final conclusions about feasibility can be made.

According to the original proposal, live observations were supposed to be carried out by the trained observers. However, this was not possible to arrange since no slaughter plant did allow this.

9.6 CONCLUSIONS

The proposed method is a promising method for assessment of behaviours related to fear and injury during driving of cattle into the stun box. However, final conclusions about feasibility of this method performed at unloading can not be made until the effect of an increased training on inter-observer agreement

10

CATTLE HEALTH STATUS

E. Canali, H.R. Whay and K.A. Leach

10.1 SUMMARY

Disease and mortality among dairy and beef cattle is a problem both in terms of welfare and economic loss. Welfare assessment protocols designed for the assessment of individual farms may take into account parameters which directly or indirectly reflect the health of dairy and beef cattle. Data regarding health are sometimes recorded on a routine basis, yet may not be readily available, or may be recorded in variable formats. A review of literature was carried out to determine the most suitable and feasible measures of health for dairy cows, fattening bulls and veal calves. A first table of possible health status measures, listing the validity, and feasibility of each measure was produced. These measures were discussed by a panel of veterinary or scientific experts in cattle health, Some measures were deleted due being less common symptoms or needing a specific clinical expertise for diagnosis. Sampling methods for measures obtained by observation were considered, and it was decided that, after using a sampling strategy which covers all buildings and production groups, measures should be summarised at a farm level. The resulting protocol covers fifteen potential health status measures. Simple measure descriptions are given for: Coughing, Sneezing, Increased respiratory rate, Nasal discharge, Ocular discharge, Diarrhoea, Mastitis, Vulvar discharge, Mortality, Bloated rumen, Tail necrosis, Dystocia, Anaemia, Downer cows, and Culling rate.The suitability of Dystocia, Downer Cows and Culling rate as measures is limited at present by the number of farms where these data are recorded. Assessing an adequate sample of animals will be time-consuming on large farms.

10.2 INTRODUCTION

The objective of this task was to put forward a proposal for the standardisation of some signs of health status in cattle. Diseases are important welfare problems. Effective health care therefore requires that cattle be kept in appropriate environments. Preventive measures, for example good hygiene and appropriate vaccination regimes, can help avoid infection of herds. Many diseases are multi-factorial. Their development may depend on the husbandry conditions of the cattle. Effective health care therefore requires that cattle are kept in environments which do not cause stress and reduced immunocompetence (SCAHAW report, 2001). Some health measures can be assessed by observations on the animals, but others can only be assessed if records are kept on the farm. There are limited publications relating to welfare assessment schemes looking at the health status parameters and how they are assessed. Clinical signs are often used but standardization of clinical signs is very rare and they are used mainly for veterinary diagnostic purposes. Therefore literature on possible health measures was reviewed, and discussions were held to decide on the most suitable and feasible measures to include in protocols to assess the health status of dairy cows, fattening bulls, and veal calves, before drawing up a protocol of recommended measures.

10.3 METHODS

Initially a review of the available literature on cattle health measures and clinical signs used in welfare assessment was carried out. The group met on one occasion in Milan (17–18 January 2006). Further work was carried out by each researcher separately and discussed via email. Drafts of the deliverables were circulated between the members. From the work carried out by each researcher it was decided that the proposal should focus on health related data just at farm level. Initial discussion focussed on the need for a tabulation of possible health status measures for cattle to ease the understanding of the variety of possible measures. According to the available literature and personal experience in preparing and using monitoring systems incorporating clinical signs, the partners prepared a first table of possible health status measures; the validity, and feasibility of each measure were listed. Although other criteria could have been selected, the ones described above appeared to be a reasonable method for ranking the measures. The second step was to discuss this table with a group of experts. The members of the group were chosen for their experience in this area (veterinarians and animal science cattle experts) in two meetings in Milan. Some measures were deleted due to the fact of being less common symptoms or needing a specific clinical expertise for diagnosis.

After the choice of health measures, particular attention was paid to the practical sampling strategy, which should suit different production systems, different concentrations of animals per group and animals with little habituation to human proximity (i.e. fattening bulls). From references and discussions with other members of the Welfare Quality project, it has been decided to record the numbers of animals which present signs as a proportion of the total number of animals in the pen or in the herd. The proposal mentions particular considerations for sampling methods for dairy cows, fattening bulls and veal calves.

10.4 RESULTS AND DISCUSSION

10.4.1 Literature Review

Respiratory and digestive disorders are quite common in beef and dairy cattle especially in young animals (Radostits et al., 1999; EFSA. 2006). Anaemia can be found in veal calves, and could be detected checking the haemoglobin levels of the animals (EFSA, 2006.)

Diarrhoea is caused by dietary factors or by infections due to viruses, bacteria or parasites. Enteritis is clinically recognized by the observation of faeces with a looser consistency than normal. Colour as well as smell of the faeces might be affected. Respiratory disorders can be caused by infection (viral, bacterial), parasites and environmental causes (i.e. ammonia,) (Radostits et al., 1999). The signs usually found are fever, nasal discharge, and coughing (Radostits et al., 1999).

Mastitis and metabolic disorders are quite common in dairy cows. Information on these pathologies can be achieved by different methods such as performing direct observations on the cows (i.e. when they are in the milking parlour), looking at the farm records (or asking the farmer questions) or looking at the somatic cell count data.

Tail tip inflammation has been reported in intensive fattening bull units with close confinement since the 1970s (SCAHAW report, 2001). The lesion occurs more often in young bulls on slatted floors while a lower frequency has also been reported in tethered bulls and sporadically in heifers kept in the same housing system as fattening bulls. This lesion is caused in most cases by traumatic injuries (tail tramping) which subsequently become infected. The lesion usually begins at the tip of the tail with a typical inflammatory reaction that gradually extends upwards. When the incidence of tail tip inflammation is high this can lead to economic loss due to reduced weight gain, and death losses due to pyaemia, and veterinary costs (SCAHAW report, 2001).

10.4.2 Measures Chosen

The measures chosen for inclusion in the protocol are shown in Table 10.1, along with the animal group for which they are relevant and their ranking in terms of suitability.

Equipment required:

- Appropriate disposable clothing and footwear
- Recording sheets, clipboard and supply of pens

TABLE 10.1 Summary of the health measures considered for cattle.

Clinical measure/record	Animals for which measure is appropriate	How to assess	Ranking
Coughing	DC / FB/ VC	Percentage on total number of animals	High
Sneezing	DC / FB/ VC	Percentage on total number of animals	High
Nasal discharge	DC / FB/ VC	Percentage on total number of animals	High
Increased respiratory rate	DC / FB/ VC	Percentage on total number of animals	High
Ocular discharge	DC / FB/ VC	Percentage on total number of animals	High
Diarrhoea	DC / FB/ VC	Percentage on total number of animals	High
Bloated rumen	VC	Percentage on total number of animals	High
Tail necrosis	FB/ VC	Percentage on total number of animals	High
Metritis	DC	Percentage on total number of animals	High
Mastitis	DC	Farm records on somatic cell counts	High
Dystocia	DC	Farm records	Medium
Downer cows	DC	Farm records	Medium
Culling rate (for accident or serious disease)	DC	Farm records	Medium
Life expectancy	DC	Farm records	Low
Mortality	DC / FB/ VC	Farm records	High

Notes: DC dairy cows; FB fattening bulls; VC veal calves.

- Spray marker
- Buckets and disinfectant for boots

10.4.3 Taking the Measurement

Precisely how each measurement should be taken will vary from parameter to parameter, however the majority are assessed by counting the number of clinical signs on the total number of animals in the pen (fattening bulls, veal calves) or in the production group (dairy cows).

The strategy to observe clinical signs varies according to the categories of animals and size of pen and groups:

- There should be no need for the inspector to enter small pens of veal calves (i.e. containing less than 10 animals). It could be possible to check to pens simultaneously (5 or less calves).
- For larger pens, the inspector will need enter the pen and walk through slowly from one side to the other.
- For fattening bulls animals should be checked without entering the pens, but the chosen pens should offer high visibility.

For dairy cows the inspector should spend 15 minutes inside the structure (however more time may be required if there are a large numbers of animals in the herd).

On entering the housing area:

- Walk slowly one step per second
- Avoid abrupt movement

10.4.4 Sampling Method

For the majority of parameters it is necessary to consider the animals on an individual basis. The sampling strategy should include samples of all different housing types on the farm. Sampling should be done across the different stages of production for fattening bulls and veal calves. All production groups should be sampled for lactating dairy cows.

10.4.5 General Sampling Notes

Dairy cows

- Cows must be sampled at every production stage.
- If the cows are divided into production groups each group should be inspected and productive stage should identified.
- If there are 15 or fewer animals in the herd, all should be sampled.
- If there are more than 15 animals in the herd, a proportion should be assessed according to Appendix.
- It is important to check a proportion of animals which approach the inspector and also a proportion of animals which avoid him/her and remain far away. The inspector will walk slowly through the herd. Cows that have been checked should be identified using stock marker spray to avoid double counting.

Veal calves and fattening bulls:

- A minimum of 4 pens per production stage (beginning, middle and end of production stage) per building is suggested, sample size on farms of different total populations should be subject to discussion at the full farm protocol stage.
- If there are different systems of buildings, all types must be sampled.
- Where there are identical buildings/systems for a production stage, these should be sampled representatively (e.g. 50% from each of two buildings).
- Where there are multiple non-identical buildings/systems for a production stage, a representative sample from each should be taken relative to the number of animals.
- It is important to assess pens in all the different part of the buildings such near door, in the middle and near fans.

- Where pens contain 15 animals or fewer, all subjects within should be sampled.
- The assessor need not enter pens with small numbers of veal calves if these are easily viewed from the passageway.
- In pens/paddocks containing over 50 calves, a proportion of the animals should be assessed according the appendix 1: Inspectors should assess calves in each part of the pen and not only those animals which approach him/her Animals that have been checked should be identified using stock marker spray to avoid double counting.
- Fattening bulls should always be assessed without entering the pen.

10.4.6 Specific Procedures for Assessing the Measures

Respiratory Problems

Respiratory problems are more common in beef cattle than in dairy cows and they can impair animal health and welfare. Coughing, sneezing (not so common in cattle), nasal and ocular discharge and increased respiratory rate (varying degrees of breathing difficulty and noise) are useful signs. Particular attention should be paid to bulls at the beginning of the fattening cycle. The method described to assess these signs is suitable for veal and fattening bulls, and can be used for cows, however they are not often observed showing signs of respiratory distress. Due to the difference in rearing systems for these categories of animals, separate descriptions of assessment are reported below.

Methods

Veal calves and fattening bulls
Coughing, sneezing:

- The inspector should spend 5 minutes per pen listening to and counting the number of coughs and sneezes. For each pen, the occurrences of each event will be counted together with the number of animals present in the pen. It could be possible to check two pens simultaneously (5 or less calves). If one or few animals show the majority of coughing and sneezing the inspector will note this.

Ocular discharge, nasal discharge and increased respiratory rate:

- For each pen, the occurrences of each clinical sign will be counted together with the number of animals present in the pen.
- There should be no need for the inspector to enter small pens of veal calves (i.e. containing 5 or less calves).
- For larger pens, the inspector will need enter the pen and walk through slowly from one side to the other assessing a proportion of animals according the sampling procedure.

Dairy cows
Coughing, sneezing:

- The inspector should spend 15 minutes inside the structure (however more time may be required if there are a large number of animals in the herd) listening to and counting the number of coughs and sneezes. If the cows are divided into production groups each group should be inspected.

Ocular discharge, nasal discharge and increased respiratory rate:

- For each group, the occurrences of ocular discharge, nasal discharge and increased respiratory rate will be counted together with the number of animals present in the group. The inspector will walk slowly through the herd assessing a proportion of animals according to the procedure described in the sampling strategy section.

Enteric Problems: Diarrhoea

Methods

- Diarrhoea is commonly assessed looking at the consistency of the faeces through the use of a faecal score. Many similar faecal scores are reported in literature as a diagnostic tool. However in order to be used in a welfare scheme we think it could be better to use a very simple such as a two point faecal score: 1=diarrhoea, loose faeces with reduced solid matter; and 2=severe diarrhoea, aqueous faeces with markedly reduced or little solid matter, blood. However, it is necessary to know the diet fed to the animals. In veal calves faeces are commonly liquid but presence in the pen of yellow faeces with a characteristic odour should be recorded with a note even in absence of animals which shows clinical signs.

Veal calves and fattening bulls

- The inspector should spend 5 minutes per pen. For each pen, the occurrence of subjects with signs of scouring (animals with loose faecal material around the anal region, faecal staining) will be counted together with the number of animals present in the pen. There should be no need for the inspector to enter small pens of veal calves (i.e. containing less than 10 animals). It may be possible to simultaneously observe two small pens for veal calves (5 calves per pen). For larger pens, the inspector will need to enter the pen and walk through slowly assessing a proportion of animals according the instructions in the sampling strategy section.

As enteritis is very common in young animals particular attention should be paid to the pens which host calves of less then 1 month old. For fattening bulls, animals should be checked without entering the pens.

Dairy cows

- The inspector should spend 15 minutes inside the structure (however more time may be required if there are a large number of animals in the herd). If the cows are divided into production groups each group should be inspected. For each group, the occurrences of subjects with abnormal consistency of the faeces or scouring signs will be counted together with the number of animals present in the group. The inspector will walk slowly through the herd assessing a proportion of animals according the procedure described in the sampling strategy section.

Enteric Problems: Bloated Rumen

Veal calves

- For each pen, each animal that shows noticeable signs of rumen distension will be counted together with the number of animals present in the pen. Sampling strategy should be as for the above clinical signs.

Enteric Problems: Anaemia

Veal calves

- Currently the most effective method is to use farm records (haemoglobin levels). Therefore in the future farms who want to enter the scheme should keep the relevant records.

Reproductive Problems (Dairy cows): Mastitis

Methods

Three different methods to assess the occurrence of mastitis on the farm can be used: performing direct observations (checking the udder) on the cows when they are in the milking parlour, looking at the farm records (or asking the farmer questions) or looking at the somatic cells counts.

- Currently the most effective method is to use somatic cell counts therefore in the future farms who want to enter the scheme should keep the relevant records.

Reproductive Problems (Dairy cows): Metritis (Vulvar Discharge)

Methods

- The inspector should spend 15 minutes inside the structure (however, more time may be required if there are a large number of animals in the herd). If the cows are divided into production groups each group should be inspected. For each group, the number of subjects with abnormal vulvar discharge will be counted together with the number of animals present in the group. The inspector will walk slowly through the herd assessing a proportion of animals as described above in the sampling strategy section.

Reproductive Problems (Dairy cows): Dystocia

Methods

- Currently the most effective method is to use farm records regarding the cows which could not have calved successfully without assistance. In the future farms who want to enter the scheme should keep the relevant records.

Other Clinical Conditions: Downer Cows (Metabolic Disorders in Dairy Cows)

Methods

- Currently the most effective method is to use farm records, In the future farms who want to enter the scheme should keep the relevant records.

Other Clinical Conditions: Tail Necrosis (Veal Calves and Fattening Bulls)

Methods

- The inspector should spend 5 minutes per pen. For each pen, the occurrence of subjects with signs of necrosis of tip of the tail will be counted together with the number of animals present in the pen. There should be no need for the inspector to enter small pens of veal calves (i.e. containing less than 10 animals). It may be possible to simultaneously observe two small pens for veal calves (5 calves per pen). For larger pens, the inspector will need enter in the pen and walk through slowly assessing a proportion of animals according to the description in the sampling strategy section.

Other Data: Mortality

Dairy cows, veal calves, fattening bulls

Methods

- Currently the most effective method is to use farm records, In the future farms who want to enter the scheme should keep the relevant records.

Life expectancy (dairy cow)

Methods

Currently the most effective method is to use farm records such as years of life, number of lactations or number of calvings. In the future farms who want to enter the scheme should keep the relevant records.

Culling rate (dairy cow)

Methods

Currently the most effective method is to use farm records, In the future farms who want to enter the scheme should keep the relevant records. (However culling will be noted only if due to severe accident or disease and inspector will not record culling for productive reasons such as low fertility or low productivity).

Examples

Photographic or video examples are not presented here but will be prepared in line with the requirements of the WQ monitoring scheme, and after decisions have been made on the practicality and usefulness of the measures described above.

Type of Data

Data are cardinal: each clinical sign is expressed in relation to the total number of animals in the pen or in the herd.

The percentage may be expressed at pen level or at farm level.

We will not use ordinal scale for the clinical signs except for diarrhoea as we think that to give a score which expresses the severity of sign (i.e. 0 no nasal discharge, 1 mild =non purulent discharge, 2=severe purulent discharge) will need not only a longer training but also specific preparation and education of the observers (veterinarians).

Validity of the Measure

The measures have not been validated within this project. However it has been recognized that disease impairs welfare (see references such as last EFSA reports on calves and beef cattle reported in the reference section).

Reliability of the Measure

No reliability testing has been carried out in this study, as the proposal was only for standardisation. Reliability testing could be carried out within training for SP2.

Feasibility of the Measures

The measures are feasible; however we do have concerns about:

- the amount of time necessary for adequate sampling, particularly on large farms with many pens or groups of animals and systems used (i.e. dairy cow tied or loose systems);
- fattening bulls: due to the danger of entering into the pen there is the necessity to check them from the outside and it could be possible that some pens can not be checked clearly from the outside (no longer a random sample of the pens, a biased choice)
- some parameters such as vulvar discharge will need to be checked with the farmer as it could be a problem to distinguish a real pathological sign from normal discharge after parturition;
- data collected from farm records (such culling, mortality) require that the farmers are already accustomed to registering these data properly.

It could be possible that some clinical signs or mortality may be more prevalent at certain periods of the year. However, none of the parameters seems to be influenced by diurnal factors, and so can be recorded at any time of day. However observations should be performed in a quiet period.

Attention should be paid to the productive phase and physiological state due to the fact that some diseases are more prevalent at certain times (i.e. downer cow after parturition, respiratory and enteric problems in first month of life in calves). This point is very important especially if the goal of the final protocol is not only a certification but also the possible check of risk factors for animal welfare and suggestions for the farmer.

Dairy cows, fattening bulls and veal calves should be assessed, although some parameters are specific to cows only (i.e. mastitis, vulvar discharge) or fattening bulls and veal calves (i.e. anaemia, tail necrosis).

The training period should be quite rapid.

Rank of the Measure (1=low to 4=high) (See Table 10.1)

We rank some of the measures as highly suitable (rank 4) for the final monitoring system. We feel that some of them can be included in the final protocol as they are valid and can be recorded rapidly. The measures ranked less then 4 are valid but need a system of keeping data that is not present at the moment in the majority of the farms. Life expectancy has been scored 1 since, although for dairy cows this could be a welfare parameter, nowadays culling is more linked to productivity than other causes.

10.5 CONCLUSIONS

The measures selected and ranked as most suitable for the final monitoring system were Coughing, Sneezing, Increased respiratory rate, Nasal discharge, Ocular discharge, Diarrhoea, Mastitis, Vulvar discharge, Mortality, Bloated rumen (veal calves) and Tail necrosis (veal calves and fattening bulls). The selected dairy cow measures of Dystocia, Anaemia, Downer cows, and Culling rate were considered slightly less suitable since the assessment, although valid, relies on records rather than animal observation, and few farms keep these records at present. Life expectancy for dairy cows was given a low suitability since in modern dairy systems it is more likely to be related to productivity than to health. There are some concerns about the time needed for adequate sampling on large farms, and bias may be introduced into some of the measures on fattening bulls by the need to observe without entering the pen.

11

ASSESSMENT OF STUN QUALITY IN CATTLE

R. Westin, A. Velarde, A. Dalmau and B. Algers

11.1 SUMMARY

An effective stunning that induce instantaneous insensibility is of great concern to assure a high standard of animal welfare. A method for assessment of stun quality has been tested and evaluated and the general conclusion is that the presented method is a good candidate for measuring stun quality in cattle.

Two hundred animals, of which at least 80 consist of bulls, are individually monitored from stunning until onset of bleeding. The symptoms displayed by the animal during this time period are marked in a stun audit protocol. According to the symptoms displayed, the stun of each animal is identified as good, poor or undefined.

The symptoms indicating the stun quality level are listed as follows.

- Good stun – The animal shows no signs of eye movements and has dilated pupils, fixed in a staring gaze.
- Poor stun – The animal show one or several of the following symptoms: corneal reflex, spontaneous blinking, righting reflex and respiration.
- Undefined stun – The animal show eyeball rotation up to sticking, nystagmus, gasping/groaning or excessive kicking in combination with eyeball rotation, nystagmus or gasping/groaning.

In addition to displayed symptoms, the animal type of each individual (bull, cow, heifer, steer, calf) is noted in the protocol.

A sum-up of all animals displaying symptoms in each category will finally give:

- the number or percentage of animals deeply stunned at first attempt;
- the number or percentage of animals poorly stunned at first attempt;
- the number or percentage of animals with an undefined stunning.

The proportion of bulls deeply/poorly stunned compared to non-bulls should also be calculated since the stunning of bulls is identified as a problem area.

11.2 INTRODUCTION

An effective stunning that induce instantaneous insensibility is of great concern to assure a high standard of animal welfare. Captive bolt stunning is the most common method to stun cattle at slaughter in Europe (Eikelenboom, 1983). However, misplacement of the captive bolt or use of incorrect cartridge can easily lead to a poor stun. Poor maintenance of the captive bolt equipment is also a common reason to poor stunning (Grandin, 1996).

A method for assessment of stunning effectiveness in cattle has recently been developed by Prof. Bo Algers and Sophie Atkinson at the Swedish University of Agricultural Sciences. This method has been used for surveys of stun quality in 7 different abattoirs in Sweden (Algers and Atkinson 2006). According to this method, stun quality is assessed by monitoring the symptoms displayed of an animal from stunning until onset of bleeding. The symptoms displayed by the animal identify the stun quality as either: 1) a good or deep stun, 2) a poor stun, 3) a stun that fits into neither deep nor poor stunning categories due to symptoms not understood for its relationship to stun quality. The symptoms indicating stun quality are described in Table 11.1.

TABLE 11.1 Symptoms that indicate stun quality level.

Stun quality	Symptoms
Deep or good stun	• Dilated pupils fixed in a staring gaze • No eye ball rotation up to sticking • Minimal kicking and reaction to sticking procedures
Poor stun	• Corneal reflex • Spontaneous blinking • Full or partial eye ball rotation up to sticking • Breathing/respirations • Righting reflex while hanging on the rail
Undefined stun quality but separating stun from deep stun quality or poor stun quality symptoms.	• Eye ball rotation up to sticking only with no other symptoms • Nystagmus • Gasping, groaning • Excessive kicking or struggling at sticking in combination with above symptoms in this category.

11.3 METHODS

During a two day visit to two different abattoirs cattle were recorded during stunning and sticking. From the collected material a film, exemplifying different behaviours displayed at stunning of cattle was made. 6 observers (2 from Spain, 4 from Sweden) where than trained in monitoring stunning effectiveness by reading written instructions and watching the film. After training, monitoring of stunning where carried out for 33 animals on video recordings for two days in a row. During monitoring each cattle was observed individually from stunning until onset of sticking. Symptoms indicating stun quality (Table 11.1) were marked in a stun audit protocol (Appendix 1). (Definitions of these symptoms are found in Chapter 3.1.) In addition to the listed behaviours, the animal type (bull, heifer, steer, cow, calf) was described for each animal monitored.

Kappa coefficients for inter-observer reliability were calculated by using the Attribute Agreement Analysis in Minitab (MINITAB® Release 14.1, Minitab Inc, 2003). Kappa coefficients (Altman, 1991) were calculated separately for signs of eye movements (dilated pupils, spontaneous blinking, nystagmus or eye ball rotation up to sticking), righting reflex (yes/no) and excessive kicking (yes/no). No monitored animal displayed corneal reflex, breathing or gasping/groaning so no statistical analysis involving these symptoms were conducted.

11.4 RESULTS

Regarding signs of eye movement, all observers obtained exact agreement in 28 cases day 1 (84.9%) and 30 cases day 2 (90.9%). The overall value of kappa for agreement between all six observers were 0.79 respectively 0.86. In two cases where no exact agreement was made, at least one observer had marked that the animal displayed ‘eye rotation’. According to the definition of ‘eye rotation’ (appendix 1), this symptom is only to be noted if it lasts until sticking. In both these cases the animal’s eyes became fixed in a blank stare before sticking and should therefore not have been marked as ‘eye rotation’ in the protocol. Two other cases involved the symptom ‘nystagmus’. In one case one observer noted ‘nystagmus’ when the others noted ‘spontaneous blinking’ and in the other case ‘eye rotation’. Only one animal was monitored with more than two different answers between the observers. This animal was re-shot shortly after falling out on the crate and the observers had only a few seconds to monitor the behaviour.

The number of exact agreements for ‘righting reflex’ were 17 day 1(51.5%) and 16 day 2 (48.5%). The corresponding values of kappa were 0.37 the first day, respectively 0.49 the second day. Statistic analysis of ‘excessive kicking’ show that exact agreement between all observers was done in 22 observations day 1 (66.7%) and 25 day 2 (75.8%). These figures give corresponding values of kappa of 0.28 and 0.61.

The value of kappa evaluates the strength of agreement between observers. It has a maximum value of 1.00 when agreement is perfect, a value of zero indicates no agreement better than chance, and a negative value shows that the agreement is worse than chance. According to Altman (1994) the strength of agreement for a value of kappa between 0 and 1.00 can be interpret as 'poor' if less than 0.20, 'fair' if 0.21–0.40, 'moderate' if 0.41–0.60, 'good' if 0.61–0.80 and 'very good' if 0.81–1.00. Using these intervals, the calculated kappa values indicate a very good inter-rater agreement for 'signs of eye movements', a good agreement for 'excessive kicking' and a fair agreement for 'righting reflex'.

11.5 DISCUSSION

If an animal isn't properly stunned it can regain consciousness and sensibility (Nijhoff, 1983). A conscious animal will be able to feel pain from the stunning, the shackling and from the sticking procedure. It is therefore of major welfare concern that as few animals as possible are at risk of regaining consciousness. A method for measuring stun quality in cattle can be a useful tool to enhance the animal welfare in European abattoirs. A report by Grandin (2005) shows that the enforcement of an animal welfare audit system at abattoirs in the United States, yielded large improvements in the handling and stun quality of cattle.

Some abattoirs do keep a record of the number of animals that are re-stunned (Karin Jonson, personal message). During the visits to the abattoirs it was noticed that some animals were re-shot without displaying any symptoms of a poor stun, and others were not re-stunned even if displaying symptoms of poor stunning. The number of animals stunned more than once is hence not equivalent to the number of animals insufficiently stunned. If only such a record was used for monitoring stun quality it could discredit abattoirs where the staff re-stun animals often 'just in case' to ensure a good stunning even if the animals do not display any symptoms of poor stunning and credit the abattoirs that never or seldom re-stun animals even if they are poorly stunned.

The kappa values for all studied behaviours increased during the second day of monitoring. Even for 'righting reflex' the kappa value increased in spite of one case less of exact agreement between all six observers. This reflects that an increased proportion of the observers agreed even if exact agreement between all observers were not obtained during the study period. The increasing values of kappa the second day shows that training and experience is important for this scoring method, in particular when monitoring signs of righting reflex and excessive kicking. Some of the observers were unfamiliar with stunning of cattle before attending to this study and they found it hard to separate the righting reflex and excessive kicking from the normal spinal reflexes that the stunning provokes. With more training the kappa values for these symptoms probably would be higher why a longer training period of observers should be considered.

The stun quality of stunned cattle depends on the application of the gun or captive bolt and the energy transmitted to the brain by the bullet or bolt (Eikelenboom, 1983). Stunning of bulls with heavy sculls has been identified as a problem area in stun audits (Algers and Atkinson, 2006; Grandin, 2005). It is therefore important that a large enough proportion of the studied animals consist of bulls. During the Swedish survey (Algers and Atkinson, 2006) a minimum of 200 animals where monitored at every abattoir and at least 80 of these were bulls. These amounts of animals were sufficient to detect differences in stun quality between bulls and other animals.

11.6 CONCLUSIONS

The presented method is a good candidate for measuring stun quality in cattle. A minimum of 200 animals should be scored for signs of eye movements, respiration, righting reflex and excessive kicking from stunning until onset of bleeding. 80 or more must be bulls since these are the most difficult to stun properly. A thorough training is important to ensure good inter-rater agreement.

12

RELIABILITY OF MEASURES OF AGONISTIC BEHAVIOUR IN DAIRY AND BEEF CATTLE

S. Laister, N. Brörkens, S. Lolli, D. Zucca, U. Knierim, M. Minero, E. Canali and C. Winckler

12.1 SUMMARY

Agonistic encounters can be regarded as normal behaviour in cattle. An increased incidence of agonistic behaviours may indicate unpleasant or stressful situations or lead to injuries. Therefore agonistic social behaviours have been regarded as potential measures to be included in an on-farm welfare monitoring scheme.

It was the aim of this study to investigate agonistic social behaviours (head butt without displacement, displacement, chasing, fighting, chasing up and the total number of these agonistic behaviours) as on-farm measures in dairy cows and fattening bulls with regard to feasibility, inter-observer reliability and intra-farm variability (consistency). Continuous behaviour sampling was carried out on three days for 4 h each on 43 dairy (31 loose housing systems, 12 tie stalls) and 20 beef farms (10 deep litter, 10 fully slatted floors) in Austria, Germany and Italy. In beef farms, three weight classes were defined (200–350 kg, 350–550 kg, 550 kg). Farm visits took place at approximately 60 and 180 days (tie stalls: 60 and 120 days) after the first visit. Inter-observer reliability was tested during direct observations (6 dairy, 2 beef farms) and using video clips (n=55).

Chasing and fighting were very rarely observed in dairy and beef cattle (median <0.10 events/animal*hour). This is also the case for chasing-up with the exception of finishing beef bulls (median frequency 0.13 events/animal*hour). However, low incidences of chasing-up made inter-observer reliability testing impossible and it did not show any consistency over time (Kendall's W=0.24). However, the behaviours chasing, fighting and chasing-up can be considered in a total agonistic behaviour measure.

Inter-observer reliability from direct observations was generally sufficient for head butts without displacement, displacements and total agonistic behaviours (W=0.83–0.97

depending on number of observers). Recording from video clips also led to acceptable levels of agreement between observers (W=0.83, 0.85 and 0.87 respectively).

In loose housed dairy cattle, consistency over time was higher for displacements than for headbutts without displacements (2h observations: W=0.75 vs. 0.70) and was also acceptable for all agonistic measures lumped together (W=0.74). However, none of the potential measures for dairy cattle in tie stalls proved to be consistent over time. In beef bulls, consistency over time increased when all agonistic behaviours were lumped together. The same was true when the weight classes medium and finishing bulls were pooled. Acceptable levels of consistency were then reached (W=0.74).

In conclusion, with regard to inter-observer repeatability and intra-farm consistency we suggest to use total agonistic behaviours as a measure of agonistic social behaviour in loose housed dairy cattle on-farm welfare assessment. Additionally, head butts without displacements and displacements may be differentiated. In beef bulls, it is recommended to differentiate between two weight classes (200–350 kg; >350 kg) and to also use as a measure total agonistic behaviours; head butts without displacement may be differentiated in beef bulls as well.

12.2 INTRODUCTION

As farm animal species live in groups or herds, non-agonistic and agonistic interactions between animals contribute to establishing and maintaining the social structure. Group size (Kondo et al., 1989), housing system and dimensions of the housing system (Wierenga, 1984), as well as management strategies (Wierenga and Hopster, 1982; Knierim, 1999; Bøe and Færevik, 2003) influence the occurrence and quality of agonistic interactions. A certain, yet mostly unknown, level of agonistic encounters can therefore be regarded as normal. An increased incidence of agonistic behaviours, however, may indicate unpleasant or stressful situations as it has been shown by e.g. mixing piglets at weaning (Otten et al., 1997), mixing beef bulls for finishing (Mounier et al., 2005) or introducing new individuals in a resident group (Albright and Arave, 1997; Coppedge et al., 1997). Furthermore, in horned cows the frequency of agonistic interactions is positively correlated with the occurrence of skin injuries (Menke et al., 1999) and it is likely that also in dehorned cows aggressive interactions result in less obvious lesions such as haematomas. Therefore agonistic social behaviours have been regarded as potential measures to be included in an on-farm welfare monitoring scheme.

Although agonistic social behaviour measures have been suggested for (Whay et al., 2003; Winckler et al., 2003) or applied in on-farm welfare assessment protocols (Capdeville and Veissier, 2001), relatively little is known about the minimum duration or the time frame of observations in order to get a representative picture of a given farm. Pilot studies in dairy herds have shown that agonistic interactions can be reliably recorded during the first hours after feeding, showing the highest inter-day repeatability for this period of the day. Moreover, it has been recommended to record interactions involving physical contact only (Winckler et al., 2002). There is, however, almost no information available on the reliability of social behaviour recordings in fattening cattle.

The objective of this study was to investigate different agonistic social behaviours such as head butts without displacements, displacements, fighting or chasing as on-farm measures in dairy cows and fattening bulls with regard to feasibility, inter-observer reliability and short- to long-term intra-farm variability (consistency). With respect to consistency, the question is addressed how representative single recordings are, considering that certain changes due to seasonal effects and management influences (e.g. regrouping, purchase of new animals) are to be expected.

12.3 METHODS

12.3.1 Inter-observer Reliability Testing (IORT)

IORT of agonistic behavioural measures was carried out based on (1) direct observations on six dairy and two beef farms and (2) recordings from video clips. For the investigation of inter-observer reliability no distinction was made between dairy cows and beef bulls as behaviour patterns are similar in both categories of cattle.

On-farm IORT

On-farm IORT was carried out three times at different stages of the project with different numbers of observers being present at each date (Table 12.1). The first testing took place after two days of initial training before on-farm data collection started. The training included discussions, video and on-farm training. The second IORT was about 50 days after observers A and B had started on-farm data collection; two additional observers (C and D) participated, who had received about half a day of training. The final on-farm IORT took place after data collection had been finished.

Observers were always located near to each other in the barn allowing a free view on the area observed.

Spearman rank correlation (rs) was used to test agreement within pairs of observers. Additionally Kendall's coefficient of concordance (W) was calculated for agreement between 3 and 4 observers, respectively.

TABLE 12.1 Overview of on-farm IORT.

Meeting		mm/yyyy	Number of farms visited	Sample size (pens, segments)			
				A	B	C	D
1	Germany	07/2005	4	35	35	–	–
2	Austria	09/2005	2	10	10	7	10
3	Italy	06/2006	2	15	15	15	–
		Total	8	60	60	22	10
		Dairy	6	50	50	22	10
		Beef[1]	2	10	10	–	–

Notes: [1] 07/2005; A, B, C, D = observers.

IORT Using Videos

After completion of the on-farm data collection and after the final on-farm IORT in June 2006, three trained observers (A, B and C) separately watched 55 video sequences of about 2 to 14 minutes (07:05 hours in total). The video clips had been recorded on different beef and dairy farms and contained representative situations of the behaviours that were in the scope of the study. The video observations followed the same rules as provided in the instructions for data collection.

Again Spearman rank correlations and Kendall's coefficient of concordance (W) were used.

12.3.2 Investigation of Intra-farm Consistency

In total, 43 dairy farms (20 cubicle, 11 deep litter, 12 tie stall systems; herd size 12–150 cows) and 20 beef fattening farms (10 deep litter, 10 fully slatted floor systems; animals per farm: 30–220, 5–27 bulls per pen) in Austria, Germany and Italy (only dairy) were included in the study (Table 12.2). The dairy cows belonged to different breeds with Holstein Friesian, Simmental-Fleckvieh and Brown Swiss being the most prevalent breeds. The fattening bulls were Simmental-Fleckvieh (S-FV), Limousin and S-FVxLimousin crosses. Other breeds such as Belgian Blue, Brown Swiss, Holstein Friesian, Tyrolean Grey or Charolais were also kept in small numbers.

Observations of agonistic behaviours as defined in Table 12.3 were carried out on three days at intervals of approximately 60 and 180 days after the first visit (beef cattle farms, dairy loose housing systems). In dairy farms with tie stalls, the third visit took place 120 days after the first visit.

TABLE 12.2 Overview of number and type of farms visited in each country.

	Austria	Germany	Italy	Total
DAIRY				
Cubicles	8	8	4	20
Deep litter	3	4	4	11
Tie stalls	6	6	–	12
Total	17	18	8	43
BEEF				
Fully slatted	5	5	–	10
Deep litter	5	5	–	10
Total	10	10	–	20

TABLE 12.3 List of behaviours observed and their definitions.

Head butt without displacement (HB)	Interaction involving physical contact where the actor is butting, hitting, thrusting, striking or pushing the receiver with forehead, horns or horn base with a forceful movement; the receiver does not give up its present position (no displacement, see definition below).
Displacement (DP)	*Dairy loose house systems & beef bulls*: Interaction involving physical contact where the actor is butting, hitting, thrusting, striking, pushing or penetrating the receiver with forehead, horns, horn base or any other part of the body with a forceful movement and as a result the receiver gives up its position (walking away for at least half an animal-length or stepping aside for at least one animal-width). 'Penetrating' is defined as an animal shoving itself between two other animals or between an animal and barn equipment (e.g. at feeding rack, at water trough, in cubicle). If after a displacement neighbouring animals also leave their feeding places but physical contact as described above is not involved, this reaction is not recorded as displacement. *Dairy tie stall systems*: Interaction involving physical contact where the actor is butting, hitting, thrusting, striking or pushing the receiver with forehead, horns, horn base or any other part of the body with a forceful movement and as a result the receiver is stepping at least one step aside or moving the head away from the drinker where it has just been drinking.
Chasing (CH)	The actor makes another animal flee by following fast or running behind it, sometimes also using threats like jerky head movements. Chasing is only recorded if it follows an interaction with physical contact. If, however, chasing occurs in the context of fighting then it is not counted separately. Chasing was not applied in tie stalls.
Fighting (FI)	Two contestants vigorously pushing their heads (foreheads, horn bases and/or horns) against each other while stemming their feet into the ground in sawbuck position and both exerting force against each other. Other agonistic interactions (head butt, displacement, chasing) are not recorded additionally as long as they are part of the fighting sequence. A new bout starts if the same animals restart fighting after more than 10 seconds or if the fighting partner changes. Fighting was not applied in tie stalls.
Chasing-up (CU)	The actor uses forceful physical contact (e.g. butting, pushing, shoving) against a lying animal which makes the receiver rise.
Total agonistic behaviour (AGO)	Total agonistic behaviour was calculated as the sum of all agonistic behaviours defined above: Loose housed dairy cows & beef bulls: AGO = HB+DP+CH+FI+CU Dairy cows in tie stalls: AGO = HB+DP+CU

Data Recording on Dairy Farms

On each farm the lactating dairy cows were observed excluding separate groups of dry or periparturient cows, as well as cows in hospital pens. Behaviour performed by or with animals in heat was not recorded.

Behaviour was recorded using continuous behaviour sampling and observations lasted for 4 h after morning feeding (or after feed had been pushed up). The observer was positioned on the feed bunk on an elevated observation chair.

In herds larger than 25 cows, the observations were carried out in segments of the barn which were expected to contain on average not more than 25 cows per segment. These segments covered all accessible areas (lying areas, feeding places, concentrate feeders, outdoor loafing areas etc.). The duration of continuous observations within each segment was adjusted to the number of segments so that each segment was observed at least once per two hours (minimum observation period 10 min) or at least twice within four hours respectively.

The number of animals which were feeding, standing/walking and lying within the segment was recorded at the beginning and at the end of each observation period. Data were then analysed as the mean number of events per animal and hour, taking the absolute frequency of behaviours, the duration of observations per pen/segment and the average number of animals in the pen/segment during the observation into account.

From all values obtained on segment level the mean incidence at herd level was calculated (occurrence of behaviour/animal*hour). Spearman rank correlations and Kendall's coefficient of concordance were used to test consistency between visits (1–60–180/120).

Data Recording on Beef Farms

Three weight classes were defined in line with the literature and common farming practice:

- initial fattening period (I): 200–350 kg
- medium fattening period (M): 350–550 kg
- finishing fattening period (F) ≥550 kg

Pens holding less than three animals were excluded from observations.

Behaviour was recorded using continuous behaviour sampling and observations lasted for 4 h after morning feeding (or after feed had been pushed up). The observer was positioned on the feed bunk on an elevated observation chair.

All weight classes present were observed for equal periods within each observation hour and each pen was observed at least twice during the 4 hour period. Pens with more than 25 bulls were divided into 2 or more segments (see dairy cattle).

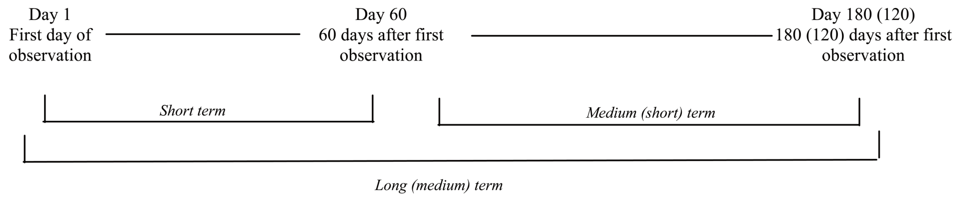

FIGURE 12.1 Schedule for on-farm observations.

Data processing within each weight class followed the same rules as for dairy cattle.

12.3.3 Decision on Measures

We followed a stepwise approach in deciding on the usefulness of the behavioural measures for on-farm welfare assessment protocols.

1. Median on-farm incidence had to exceed 0.10 events/animal*hour to allow for reliable and feasible recording as well as differentiation between farms.
2. In terms of inter-observer reliability (IOR), correlation between observers had to be r_s or Kendall's W≥0.70.
3. In terms of intra-farm consistency, again for the remaining measures correlation coefficients between results from different observation days had to be Kendall's W≥0.70.
4. For reliable and consistent measures, results from reduced observation times were checked again for intra-farm consistency.

12.4 RESULTS

12.4.1 Incidences of Agonistic Social Behaviours

In dairy as well as in beef cattle head butts without displacement (HB) were the most frequent agonistic behaviour observed (Table 12.4). In dairy loose house systems it occurred more than twice as often as in tie stall systems (median 0.54/animal*hour vs. 0.24/animal*hour). In beef bulls head butts without displacements were most frequent in medium and rarest in initial fattening bulls.

TABLE 12.4 Descriptive measures (median, mean, minimum (Min) and maximum (Max) values, standard deviation of the mean (SD) and variance (Var)) of incidences of agonistic social behaviours in dairy cows and beef bulls (events/animal*hour)

Behaviour	Housing system	Day Weight	Median	Mean	Min	Max	SD	Var	N
Total agonistic behaviour (AGO)	DAIRY Loose housing	1	*1.11*	1.43	.43	4.53	1.01	1.02	31
		60	*1.10*	1.30	.21	4.29	.81	.65	31
		180	*1.28*	1.45	.25	3.73	.95	.90	31
		overall	*1.11*	1.40	.21	4.53	.92	.84	93
	DAIRY Tie stalls	1	*.35*	.39	.00	.84	.23	.05	12
		60	*.27*	.33	.05	.92	.24	.06	12
		120	*.16*	.22	.00	.69	.22	.05	12
		overall	*.29*	.31	.00	.92	.24	.06	36
	BEEF	I	*3.32*	3.50	1.03	8.14	1.73	2.98	56
		M	*5.15*	5.62	1.00	14.81	3.29	1.85	57
		F	*4.80*	5.29	.45	14.25	3.23	1.43	57
Head butts without displacement (HB)	DAIRY Loose housing	1	*.45*	.73	.03	2.99	.62	.39	31
		60	*.63*	.66	.03	1.89	.43	.19	31
		180	*.56*	.73	.07	2.13	.57	.33	31
		overall	*.54*	.70	.03	2.99	.54	.30	93
	DAIRY Tie stalls	1	*.35*	.35	.00	.75	.21	.04	12
		60	*.25*	.29	.05	.81	.22	.05	12
		120	*.13*	.20	.00	.66	.22	.05	12
		overall	*.24*	.28	.00	.81	.22	.05	36
	BEEF	I	*1.57*	1.95	.44	5.25	1.22	1.49	56
		M	*3.50*	3.58	.18	1.13	2.36	5.55	57
		F	*3.10*	3.65	.15	1.95	2.53	6.40	57
Displacements (DP)	DAIRY Loose housing	1	*.56*	.64	.16	1.72	.45	.20	31
		60	*.49*	.59	.06	2.64	.47	.22	31
		180	*.57*	.65	.06	1.71	.44	.19	31
		overall	*.51*	.63	.06	2.64	.45	.20	93
	DAIRY Tie stalls	1	*.03*	.04	.00	.09	.04	.00	12
		60	.02	.04	.00	.14	.05	.00	12
		120	.00	.01	.00	.11	.03	.00	12
		overall	.00	.03	.00	.14	.04	.00	36
	BEEF	I	*1.23*	1.41	.38	2.89	.76	.58	56
		M	*1.63*	1.89	.15	6.17	1.23	1.52	57
		F	*1.34*	1.46	.15	4.65	.90	.82	57
Chasing (CH)	DAIRY Loose housing	1	.00	.01	.00	.06	.02	.00	31
		60	.00	.01	.00	.08	.02	.00	31
		180	.00	.01	.00	.10	.02	.00	31
		overall	.00	.01	.00	.10	.02	.00	93
	BEEF	I	.00	.01	.00	.19	.03	.00	56
		M	.00	.00	.00	.11	.02	.00	57
		F	.00	.00	.00	.00	.00	.00	57

Notes: for beef cattle overall measures for the three farm visits are presented; figures in italics exceed threshold of 0.1 events/animal*hour.

In tethered dairy cows displacements (DP) were only rarely observed (median: 0 events/animal*hour), while it occurred as often as head butts without displacements in

TABLE 12.4 CONT. Descriptive measures (median, mean, minimum (Min) and maximum (Max) values, standard deviation of the mean (SD) and variance (Var)) of incidences of agonistic social behaviours in dairy cows and beef bulls (events/animal*hour)

Behaviour	Housing system	Day Weight	Median	Mean	Min	Max	SD	Var	N
Fighting (FI)	DAIRY Loose housing	1	.00	.01	.00	.08	.02	.00	31
		60	.00	.00	.00	.03	.01	.00	31
		180	.00	.01	.00	.07	.02	.00	31
		overall	.00	.01	.00	.08	.02	.00	93
	BEEF	I	.00	.04	.00	.45	.08	.01	56
		M	.00	.05	.00	.35	.09	.01	57
		F	.00	.05	.00	.56	.10	.01	57
Chasing-up (CU)	DAIRY Loose housing	1	.00	.04	.00	.51	.10	.01	31
		60	.02	.04	.00	.19	.06	.00	31
		180	.02	.04	.00	.21	.05	.00	31
		overall	.02	.04	.00	.51	.07	.01	93
	DAIRY Tie stalls	1	.00	.00	.00	.04	.01	.00	12
		60	.00	.00	.00	.00	.00	.00	12
		120	.00	.00	.00	.03	.01	.00	12
		overall	.00	.00	.00	.04	.01	.00	36
	BEEF	I	.08	.10	.00	.55	.13	.02	56
		M	.03	.10	.00	.68	.13	.02	57
		F	*.13*	.13	.00	.59	.13	.02	57

loose housed herds (median 0.51/animal*hour). The incidence of displacements in beef cattle was about half of that of head butts without displacements.

In dairy as well as in beef cattle chasing (CH), fighting (FI) and chasing up (CU) occurred less than 0.10 times per animal and hour except for CU in finishing bulls (median: 0.13 events/animal*hour).

In conclusion, measures of agonistic social behaviour that are feasible for on-farm welfare assessment are presented in Box 12.1.

12.4.2 Inter-observer Reliability

On-farm Observations

Measures of inter-observer agreement could only be calculated for the behaviours total agonistic behaviours (AGO), head butt without displacement (HB) and displacement (DP) as observations did not provide enough data for the other behaviours such as chasing (only observed once by one observer), fighting (observed only twice by different observers) or chasing up (observed only once by observers A & B).

BOX 12.1 Measures of agonistic social behaviour that are feasible for on-farm welfare assessment.

Dairy cows	*Beef bulls*
Total agonistic behaviour	Total agonistic behaviour
Head butts without displacement	Head butts without displacement
Displacements	Displacements
	Chasing up

TABLE 12.5 Inter-observer reliability in different test sessions (Spearman rank correlation coefficients, r_s, Kendall's coefficient of concordance, W) for total agonistic behaviours, head butts without displacements and displacements

Behaviour	Session		Observer pairs						Kendall's W	
			A_B	A_C	A_D	B_C	B_D	C_D	ABC	ABCD
Total agonistic behaviour (AGO)	1	r_s	.78							
		p	.000							
		n	35							
	2	r_s	.79	.76	.94	.83	.94	.95		
		p	.000	.028	.000	.010	.000	.000		
		n	10	7	10	7	10	7		
	3	r_s	.96	.94		.96				
		p	.000	.000		.000				
		n	15	15		15				
	Overall	r_s	.89	.92	.94	.96	.94	.95	.97	.84
		p	.000	.000	.000	.000	.000	.000	.000	.000
		n	60	22	10	22	10	7	22	7
Head butts without displacement (HB)	1	r_s	.70							
		p	.000							
		n	35							
	2	r_s	.83	.61	.79	.82	.79	1.00		
		p	.003	.148	.006	.023	.006	.000		
		n	10	7	10	7	10	7		
	3	r_s	.96	.88		.86				
		p	.000	.000		.000				
		n	15	15		15				
	Overall	r_s	.83	.83	.79	.87	.79	1.00	.92	.83
		p	.000	.000	.006	.000	.006	.000	.000	.000
		n	60	22	10	22	10	7	22	7
Displacements (DP)	1	r_s	.92							
		p	.000							
		n	60							
	2	r_s	.95	.81	.92	.94	.89	.67		
		p	.000	.027	<.001	.002	<.001	.102		
		n	10	7	10	7	10	7		
	3	r_s	.94	.86		.90				
		p	.000	.000		.000				
		n	15	15		15				
	Overall	r_s	.94	.83	.92	.87	.89	.67	.92	.85
		p	.000	.000	<.001	.000	<.001	.102	.000	.000
		n	60	22	10	22	10	7	22	7

Agreement on HB varied between 0.61 and 1.00 (r_s) for single pairs of observers at different test sessions (Table 12.5). Generally, agreement increased with repeated sessions. Considering the agreement across all sessions, rank correlations ranged between 0.79 and 1.00 with the lowest correlations in pairs with observers having the least experience. Kendall's coefficient of concordance for both three and four observers was above 0.80.

With regard to DP, inter-observer reliability in live observations reached high levels with rank correlations always between 0.80 and 0.95 except within the pair of less trained observers (C_D). The calculated parameter AGO reached satisfactory to good rank correlations (r_s=0.76–0.96) and showed very good overall agreement within three and within four observers respectively (Kendall's W =0.97/0.84).

Video Observations

The total frequencies of behaviours observed in 55 video clips are given in Table 12.6. As chasing, fighting and chasing up only rarely or never occurred on the videos, these measures were only included in the analysis of total agonistic behaviour.

TABLE 12.6 Frequencies of behaviours observed by different observers (A, B, C) on 55 video clips.

Behaviour	Total frequency of observed behaviours		
	Observer A	Observer B	Observer C
Head butts without displacements (HB)	177	160	119
Displacements (DP)	143	113	126
Chasing (CH)	1	0	2
Fighting (FI)	1	2	0
Chasing up (CU)	5	7	5
Total agonistic behaviour (AGO)	327	282	252

TABLE 12.7 Spearman rank correlation coefficients (r_s) and Kendall's coefficient of concordance (W) for inter-observer agreement based on video clips (n=55).

Behaviour		Observer pairs			Kendall's W
		A_B	A_C	B_C	
Head butts without displacement (HB)	r_s	.81	.72	.70	.83
	p	.000	.000	.000	.000
Displacements (DP)	r_s	.79	.84	.71	.85
	p	.000	.000	.000	.000
Total agonistic behaviour (AGO)	r_s	.87	.87	.79	.89
	p	.000	.000	.000	.000

Box 12.2 Measures of agonistic social behaviour that revealed sufficient inter-observer reliability.

Dairy cows	*Beef bulls*
Total agonistic behaviour	Total agonistic behaviour
Head butts without displacement	Head butts without displacement
Displacements	Displacements

As for live observations, acceptable IOR was found for head butts without displacement, displacements and total agonistic behaviour with Kendall's W being 0.83, 0.85 and 0.89, respectively. However, rank correlations between single pairs of observers in some cases only reached moderate levels but were always above 0.70 (Table 12.7).

In conclusion, measures of agonistic social behaviour that revealed sufficient inter-observer reliability are presented in Box 12.2.Consistency over different periods of time (intra-farm variability)

12.4.3 Consistency over Different Periods of Time (Intra-farm Variability)

Dairy Cows

In tie stalls, levels of head butts without displacements (HB) and displacements (DP) were rather inconsistent with Kendall's W=0.54 for both parameters (Table 12.8). Merging all agonistic behaviours (AGO) did not increase consistency.

In loose housing systems, levels of DP were more consistent than HB with rank correlation coefficients between farm visits being always equal or above 0.70 (HB: below 0.70). Results for the sum parameter AGO were slightly more consistent than for HB. However, the overall repeatability of head butts without displacement, displacements as well as total agonistic behaviours reached levels above the threshold (Kendall's W = 0.74–0.84).

Fattening Bulls

Regarding short-, mid- and long-term repeatability of head butts without displacement, displacements or total agonistic behaviour most rank correlation coefficients were below 0.70 (Table 12.9). Also the overall consistency in finishing bulls was always below 0.70. Kendall's W above this threshold were reached in initial and medium weight bulls as well as when no distinction between medium weight and finishing bulls was made (M&F).

TABLE 12.8 Spearman rank correlation coefficients (r_s) for short- (1_60, 60_120), mid- (60_180, 1_120) and long-term (1_180) intra-farm consistency and Kendall's coefficient of concordance (W) for total agonistic behaviours, head butts without displacements and displacements in dairy cattle.

Housing System	Behaviour		1_60	60_120	1_120	Kendall's W
DAIRY	Total agonistic behaviour	r_s	.59	.48	–.12	.54
Tie stalls	(AGO)	p	.045	.114	.704	.046
	Head butts without displacement	r_s	.41	.58	–.06	.54
	(HB)	p	.183	.047	.846	.047
	Displacements	r_s	.43	.28	.18	.54
	(DP)	p	.161	.379	.573	.051
			1_60	60_180	1_180	
DAIRY	Total agonistic behaviour	r_s	.66	.63	.66	*.76*
Loose housing	(AGO)	p	.000	.000	.000	.000
	Head butts without displacement	r_s	.65	.60	.57	*.74*
	(HB)	p	.000	<.001	<.001	.000
	Displacements	r_s	.70	.75	.85	*.84*
	(DP)	p	.000	.000	.000	.000

Notes: figures in italics exceed threshold of W=0.70.

Reliability of Recordings from Reduced Observation Time

For the behaviours head butt without displacement and displacement correlations between results derived from four hour observations and two or one hour observations were above 0.80 and highly significant (Table 12.10).

In loose housed dairy cows and beef bulls intra-farm consistency across time was recalculated comparing the 4 h observations with the first two hours (hours 1+2) and the second two hours of observation (hours 3+4; Table 2.11 and Table 12.12).

For both HB and DP as well as AGO Kendall's W for hour 1+2 generally slightly decreased compared to the total observation time. Consistency further decreased when only data from hour 3+4 were taken into account. Considering the first two hours of observation, in dairy cows the threshold level of 0.70 was reached regarding all measures. In beef cattle, this was the case for HB and DP in the initial weight category (I) and when medium and finishing bulls were taken together (M&F) whereas AGO exceeded the threshold only in M&F bulls.

In conclusion, measures of agonistic social behaviour recommended for an on-farm welfare assessment are presented in Box 12.3.

Box 12.3 Measures of agonistic social behaviour recommended for an on-farm welfare assessment.

Dairy cows (only loose housed)	*Beef bulls*
Total agonistic behaviour	Total agonistic behaviour
Head butts without displacement	Head butts without displacement
Displacement	Displacement

Table 12.9 Spearman rank correlation coefficients (r_s) for short- (1_60), mid- (60_180) and long-term (1_180) intra-farm consistency and Kendall's coefficient of concordance (W) for total agonistic behaviour, head butts without displacement, displacements and chasing up in fattening bulls of different weight classes.

Behaviour	Weight class		1_60	60_180	1_180	Kendall's W
Total agonistic behaviour (AGO)	Iinitial (I)	r_s	.76	.63	.58	*.79*
		p	.000	.005	.01	.000
	Medium (M)	r_s	.42	.60	.51	.68
		p	.083	.008	.026	.000
	Finishing (F)	r_s	.46	.67	.47	.68
		p	.055	.002	.041	.000
	M & F[1]	r_s	.49	.66	.66	*.74*
		p	.038	.003	.002	.000
Head butts without displacement (HB)	Initial (I)	r_s	.74	.53	.51	*.73*
		p	<.001	.023	.026	.000
	Medium (M)	r_s	.52	.60	.49	*.70*
		p	.027	.009	.035	.000
	Finishing (F)	r_s	.35	.74	.52	.68
		p	.152	<.0001	.021	<.001
	M & F[1]	r_s	.60	.79	.58	*.77*
		p	.008	.000	.009	.000
Displacements (DP)	Initial (I)	r_s	.45	.04	.24	.50
		p	.060	.871	.323	.044
	Medium (M)	r_s	.61	.66	.67	*.77*
		p	.008	.003	.002	.000
	Finishing (F)	r_s	.43	–.10	.18	.45
		p	.072	.704	.450	.105
	M & F[1]	r_s	.68	.69	.84	*.82*
		p	.002	.001	.000	.000
Chasing–up (CU)	Initial (I)	r_s	.19	–.31	.16	*.86*
		p	.441	.217	.523	.000
	Medium (M)	r_s	–.08	–.04	.02	.30
		p	.746	.864	.945	.620
	Finishing (F)	r_s	–.44	.05	–.05	.24
		p	.069	.854	.839	.847
	M & F[1]	r_s	–.23	–.11	.06	.27
		p	.365	.675	.819	.746

Notes: [1] M & F = all bulls > 350 kg (merging data from pens with medium and finishing weight class bulls); I=200–350 kg, M=350–550 kg, F=>550 kg; figures in italics exceed threshold of W=0.70.

TABLE 12.10 Spearman rank correlation coefficients for incidence of head butts without displacement and displacements in loose housed dairy and beef cattle between 1 and 4 hours and 2 and 4 hours of observation.

Animals	Behaviour	Day	1h_4h	p	2h_4h	p	n
Dairy (loose housed)	Head butts without displacement (HB)	1	.77	.000	.91	.000	31
		60	.75	.000	.87	.000	31
		180	.79	.000	.85	.000	31
		overall	.78	.000	.89	.000	93
	Displacements (DP)	1	.73	.000	.92	.000	31
		60	.75	.000	.80	.000	31
		180	.81	.000	.88	.000	31
		overall	.78	.000	.87	.000	93
Beef	Head butts without displacement (HB)	1	.92	<.001	.94	<.001	20
		60	.92	<.001	.87	<.001	18
		180	.96	<.001	.94	<.001	19
		overall	.94	<.001	.96	<.001	57
	Displacements (DP)	1	.82	<.001	.90	<.001	20
		60	.95	<.001	.97	<.001	18
		180	.84	<.001	.94	<.001	19
		overall	.89	<.001	.93	<.001	57

TABLE 12.11 Kendall's coefficient of concordance (W) for consistency over farm visits for total agonistic behaviour, head butts without displacement and displacements in loose housed dairy cows for the first and second two hours of observation.

Behaviour		Total observation time (4 h)	Hours 1&2	Hours 3&4
Total agonistic behaviour (AGO)	W	.76	.74	.55
	p	.000	.000	.003
Head butt without displacement (HB)	W	.74	.70	.50
	p	.000	.000	.015
Displacement (DP)	W	.84	.75	.65
	p	.000	.000	.000
	n	31	29	29

12.5 DISCUSSION

In this study the agonistic behaviours head butt without displacement (HB), displacement (DP), chasing (CH), fighting (FI), chasing-up (CU) and the total sum of agonistic behaviours (AGO) were tested with regard to inter-observer reliability and short-, mid- and long-term consistency in dairy cattle and beef bulls. We focused on interactions involving body contact; threats and avoiding were not considered because such behaviours can be less clearly defined and likely are prone to subjective interpretation of e.g. subtle movements. We did therefore not expect to reach acceptable levels of inter-observer reliability.

During repeated on-farm visits, in both dairy and beef cattle chasing and fighting were very rarely observed regardless of the housing system (median incidence 0.0 events per

TABLE 12.12 Kendall's coefficient of concordance (W) for consistency over farm visits for total agonistic behaviour, head butts without displacement and displacements in beef bulls for the first and second two hours of observation.

Behaviour	Weight class		Total observation time (4 hours)	Hour 1&2	Hour 3&4
Total agonistic behaviours (AGO)	I	W	*.79*	.66	.33
		p	.000	.001	.490
	M	W	.68	.66	.45
		p	.000	.000	.119
	F	W	.68	.63	.52
		p	.000	.002	.030
	M & F	W	*.74*	*.72*	.61
		p	.000	.000	.003
Head butts without displacement (HB)	I	W	*.73*	*.70*	.55
		p	.000	.000	.013
	M	W	*.70*	.67	.56
		p	.000	.000	.011
	F	W	.68	.60	.56
		p	<.001	.004	.011
	M & F	W	*.77*	*.72*	.63
		p	.000	.000	.001
Displacements (DP)	I	W	.50	.48	.41
		p	.044	.069	.211
	M	W	*.77*	.73	.43
		p	.000	.000	.147
	F	W	.45	.33	.38
		p	.105	.491	.305
	M & F	W	*.82*	*.77*	.48
		p	.000	.000	.069
	n		18	18	18

Notes: figures in italics exceed threshold of W=0.70.

animal and hour). Such very low incidences of behaviour make data recording and interpretation difficult and would require long-term continuous observations in order to get reliable information on these specific types of behaviour. Chasing-up was also rarely observed in dairy cows as well as in initial and medium weight bulls. Only in finishing beef bulls it occurred at a median frequency of 0.13 per animal and hour and was therefore included in further data analysis. However, inter-observer reliability testing was not possible because it did not occur during the direct observations and was too infrequent also on the video clips. It furthermore did not show any consistency over time in finishing bulls (Kendall's W=0.24). However, such rare behaviours can be further considered in on-farm assessment protocols through the parameter total agonistic behaviours (AGO).

With regard to the remaining parameters (head butts without displacement, displacements, total agonistic behaviour), inter-observer reliability for direct observations was generally sufficient (Kendall's W=0.83–0.97 depending on number of observers). For head butts without displacement, inter-observer reliability was often higher within pairs of trained observers compared to pairs with less trained observers. More importantly it could be shown in the course of the different testing sessions that experience increased agreement over time, thus underlining the impact of thorough training. The rather low agreement regarding displacements within the pair of less trained observers in their first test session (rs C_D=0.67) is in this line. Recording head butts without displacement, displacements and total agonistic behaviour from video clips also lead to acceptable levels of agreement between (then experienced) observers (Kendall's W=0.70–0.90), which nevertheless were slightly lower than those derived from live observations. This might have been caused by the two-dimensionality which makes it more difficult to estimate whether a change in position as given in the definition has really occurred (one cow width or half a cow length, respectively). Furthermore it is likely that recording from video clips increases the chance of disagreement on occurrence and forcefulness of physical contacts due to deficits in light conditions, sound and resolution.

Since 4 h observations are likely to be too time-consuming within an on-farm assessment protocol, shortening the observation time was simulated. It was shown that reducing the on-farm observations from 4 h to 2 h would not cause a major loss of information. Consistency measures discussed below therefore refer to these reduced data sets (Tables 14 and 15).

In dairy tie stall systems, headbutts without displacements as well as displacements showed very low consistencies (0.54 for both parameters). This might be due to the rather low frequencies of the behaviours (less then half the frequency of head butts per time unit in loose housing systems, almost no displacements). Merging all agonistic behaviours into the parameter AGO did not improve consistency and therefore none of the parameters can be recommended for tethered animals.

In dairy loose housing systems, the occurrence of displacements was more consistent than of headbutts without displacements. Although with increasing length of interval between farm visits one could expect more changes in herd composition or management routines and thus in the incidence of certain behaviours, no such effect on consistency measures was found. However, considering the overall repeatability, both headbutts without displacement and displacements as well as total agonistic behaviour showed acceptable consistency (Kendall's W=0.70–0.75).

Also in beef bulls, repeatability of social behaviour measures did not depend on the length of interval between farm visits. In the different weight classes, acceptable overall consistency levels were only reached for headbutts without displacement in initial fattening bulls and for displacements in medium weight bulls. Therefore, considering each weight class separately, no single behaviour parameter would meet the requirements. Pooling medium and finishing bulls (M&F) resulted in acceptable consistency in both parameters. Such pooling could furthermore improve feasibility of the recordings since live weight estimation appeared to be rather difficult in heavier bulls (about 550 kg) and groups were often heterogeneous. Yet, a distinction between younger bulls (up to 350 kg) and older ones should still be made since the younger animals are not sexually mature yet and the development of dominance relationships can be expected in bulls of the initial weight class (Bouissou, 1985; Schloeth, 1961). However, this would mean that only head butts could be included (Kendall's W: I=0.70, M&F=0.72) or that the initial fattening bulls would have to be dropped from the protocol. A feasible option for taking all weight classes and behavioural measures into account is again to use the measure total agonistic behaviour (AGO) for which the threshold for consistency across farm visits was exceeded (Kendall's W≥0.70) in initial (I) as well as in M&F weight bulls.

12.6 CONCLUSIONS

With regard to inter-observer repeatability and intra-farm consistency we suggest to use the measures total sum of agonistic behaviours, head butts without displacements and displacements in loose housed dairy cattle for on-farm welfare assessment. None of the potential measures for dairy tie stall cattle proved to be reliable as defined in this study. In beef bulls, it is recommended to differentiate between two weight classes (I= 200–350 kg; M&F= >350 kg) and to record total agonistic behaviours consisting of all single agonistic behaviours and head butts without displacement.

13

VALIDATION OF SOCIAL LICKING AS AN INDICATOR FOR POSITIVE EMOTIONS

S. Laister, A-M. Regner, K. Zenger, C. Winckler, N. Brörkens, R. Quast and U. Knierim

13.1 SUMMARY

Social licking in cattle is considered to be a socio-positive behaviour during which pleasant emotions are likely to occur. It is either performed spontaneously or after solicitation but can also follow agonistic behaviour. Psychological and physiological calming effects of social licking have been postulated. It was the goal of this study to investigate the validity of social licking as an indicator of good animal welfare with regard to heart rate measurements in both acting and receiving animals in two different housing systems.

Behavioural observations and heart rate measurements were carried out on 24 lactating Holstein-Friesian dairy cows kept in a tie stall with zero-grazing as well as on 20 lactating Fleckvieh-Simmental dairy cows kept in a sloped-floor deep litter loose housing system with an outdoor loafing area. Heart rates were measured with 10 and 20 heart rate meters (Polar Horse Trainer S810i) at a time, respectively. Licking bouts and general activity of the cows were observed for 3 h in the morning after milking and feeding by direct observations and from video recordings (tethered animals only). Using general linear mixed models, heart rates during licking were compared with the mean heart rate during 5 min before and after licking as well as with the median heart rate over the whole observation period.

In tethered animals, only actors had lower heart rates during licking compared with the period before. Spontaneous licking had a larger effect than licking after solicitation. While being licked after solicitation numerically, but non significantly reduced heart rate, it was increased when receivers were spontaneously licked. Actors licking after solicitation and receivers showed increased heart rates during licking compared to the median heart rate over the whole observation period. In loose housed cows, heart rates were reduced in actors licking spontaneously only when they were standing or feeding but increased in lying actors. Receivers showed lower heart rates only when they were licked after solicitation.

We found indications of calming effects of licking, although the findings from tied and loose housed animals are not completely consistent. Altogether it is very likely that especially voluntary licking and being licked usually increases well-being of the individual, no matter if receiver or, according to our results from tie stalls, actor. However, on herd level it is questionable whether high numbers of animals performing licking always reflect an overall better welfare compared to animals of a herd with lower licking frequencies. In the literature it is suggested that licking may also be carried out in order to reduce social tension or to reduce the own stress response to restrictive conditions. Our results are consistent with these hypotheses in that in the tie stalls more pronounced and consistent heart rate reductions were found in actors, especially in those that licked spontaneously. Moreover, the increased heart rates compared to the median heart rates my indicate that licking occurred during a stressful period or was induced by it, although this needs further investigation.

We conclude that social licking might indicate positive effects on (at least part of) the animals involved but on the basis of changes in heart rate at present cannot be regarded as a true indicator of positive emotions. Therefore, the use of social licking as a welfare measure in on-farm assessment protocols is not recommended.

13.2 INTRODUCTION

In the past, usually the assessment of animal welfare focussed on indications of impaired welfare. During the last years, however, the interest in indicators of good welfare, namely positive emotions, has increased (e.g. Knierim et al., 2001; Désiré et al., 2002). The search for such indicators is at its early beginnings and has not yet generated many candidate indicators, especially not those that are suitable for use in an on-farm welfare assessment (Winckler et al., 2003).

Social licking is often suggested as a positive candidate indicator for cattle welfare (e.g. Knierim et al., 2001; Winckler et al., 2003). It is a form of non-agonistic behaviour (Sato et al.. 1993) which is either performed spontaneously or after solicitation (Sambraus, 1969; Sato et al., 1991) but can also follow agonistic behaviour (Reinhardt, 1980). Cows receiving licks often show behavioural signs of enjoyment such as partly closing their eyes. Based on this observation, Sato et al. (1991) postulate a psychological and physiological calming effect of social licking which was also suggested by Sambraus (1969) with regard to the licking of cows in estrous by bulls. Physiological measurements, namely of heart rate of animals receiving social grooming confirmed this hypothesis in primates (Boccia et al., 1989; Aureli et al., 1999) and cows (Sato and Tarumizu, 1993). However, in all cases sample size was minimal and analysis of the data partly questionable.

It was the goal of this study to investigate the physiological effects of social licking more closely. Our question was not only how the receiver, but also how the actor may experience the licking. Moreover, we wondered about results and assumptions that social licking increases under more restrictive housing conditions (Reinhardt, 1980; Krohn, 1994; Emmerig, 2004). Krohn (1994) reports that cows kept in tie stalls showed higher frequencies of social licking than cows in loose housing which we can confirm from own unpublished observations. Therefore, we investigated the effects of licking on heart rate in two different housing systems - tie stalls and cubicles - in order to explore possible differences in the patterns of responses.

It was our basic assumption that decreases in heart rates in licked or licking animals would indicate relaxation or calming which is expected to be experienced as pleasant emotion while increases in heart rate may be due to tension or fear which likely will be experienced as unpleasant.

13.3 METHODS

13.3.1 Dairy Cows in a Tie Stall Barn

Behavioural observations and heart rate measurements were carried out for 8 days on a privately owned German farm with a tie stall barn for 23 lactating cows and zero-grazing. Measurements were taken with 10 heart rate monitors (R-R intervals; Polar Horse Trainer S810i) at a time for 3 hours each day after morning milking and feeding. Behavioural observations started twenty to thirty minutes after attaching the belts with the monitors. General activity of the cows (standing, lying, feeding or drinking) was assessed using instantaneous scan sampling every 2 minutes. All occurrences of social licking were recorded by one observer positioned on the feed table. The exact start (at first tongue stroke) and end (immediately after last tongue stroke) as well as the category of licking (Table 13.1) was later determined from video recordings taken with three stationary video cameras covering the whole barn. Data from in total 24 Holstein-Friesian cows were recorded.

Table 13.1 Definition of licking categories.

Licking categories	Definition
Spontaneous (as actor or receiver)	The licking is not overtly initiated by the receiving cow.
Solicited (as actor or receiver)	Licking following a clear request: approach of a cow in a submissive posture of head and neck, possibly touching the other cow's mouth or gently butting against her neck if the other cow does not react.
After agonistic interaction (as actor or receiver)	Licking that occurs directly after butting or fighting between the same animals involved.

13.3.2 Loose Housed Dairy Cows

The study was carried out in November/December 2005 with in total 20 Simmental-Fleckvieh dairy cows housed in a sloped floor deep litter housing system with solid concrete floors in the passages. The animals had permanent access to an outdoor loafing area and were fed indoors, with additional grass silage being offered in the outdoor loafing area. On in total 16 days, 10 (days 1–5) or 20 animals (days 6–16) were fitted with heart rate monitors (R-R intervals, Polar Horse Trainer S810i) during and shortly after the morning milking.

Behaviour observations were carried out by two observers, started approximately 10 min after the last animal had been equipped with a heart rate monitor and lasted for 3 to 4 hours. Instantaneous scan sampling at 5 min intervals was used to assess the cows' general activity (standing/walking, feeding/drinking, lying). All occurrences of social licking were recorded with the exact start and end time (for categories, see Table 13.1).

13.3.3 Data Processing and Statistical Analysis

Only licking bouts ≥10 seconds and heart rate recordings containing not more than 5% identified errors were taken into account. Heart rate curves were corrected up to two times with moderate filter power. Additionally, heart rates lower than 40 beats per minute and higher than 180 beats per minute were regarded as being unrealistic under the given conditions and excluded. Moreover, only those licking bouts were investigated where the general activity remained unchanged in the period before and during licking.

Means of heart rates during licking, from the period of 5 minutes before and after licking, and as additional reference values, medians over the time of the observation period the cow was performing the same activity as she was doing while licking or being licked (referred to as reference period) were calculated.

Statistical analysis was carried out separately for tied and loose housed animals and for both actors and receivers. Data were analysed with general linear mixed models (SPSS 12.0). Firstly, the time of measurement (before, during or after licking) was taken into account as fixed effect, with individual cow and day of measurement as random effects. In a second step, separate analyses were carried out for the different categories of licking, with time of measurement, general activity and their interaction as fixed effects. Individual cow and day of measurement were again random effects.

13.4 RESULTS

13.4.1 Dairy Cows Housed in a Tie Stall Barn

In total, 117 recorded licking bouts fulfilled the criteria mentioned above (Table 13.2). Heart rate measures for social licking after agonistic interaction were only obtained twice in receivers.

Overall Effects of Licking

Actors had lower heart rates during licking than before licking and tended to have higher heart rates compared to the median heart rate over the whole observation period (Table 13.3). After licking, heart rate did not further change in actors.

Table 13.2 Number of valid licking bouts per category in tied animals.

Licking categories		Activity	Number of licking bouts
Actor	Spontaneous	Standing	24
	Solicited	Standing	10
		Lying	4
Receiver	Spontaneous	Feeding	27
		Standing	8
		Lying	12
		Lying & Feeding	5
	Solicited	Feeding	11
		Standing	14
	After agonistic interaction	Feeding	1
		Standing	1
Total			117

Table 13.3 Effects of licking on heart rates in actors and receivers – tied cows.

		Differences between LS means (beats per minute)	F–value	Standard error	p–value
Actors	Licking vs. 5 min before	–2.4	$F_{1,52}$=16.981	.58	.000
	Licking vs. 5 min after	–.2	$F_{1,52}$=12.607	.68	.806
	Licking vs. Reference period	+1.5	$F_{1,52}$=3.684	.77	.060
Receivers	Licking vs. 5 min before	–.7	$F_{1,118}$=.646	.82	.423
	Licking vs. 5 min after	–.7	$F_{1,126}$=3.735	.77	.390
	Licking vs. Reference period	+1.6	$F_{1,125}$=3.935	.78	.049

In receivers, heart rate during licking did not significantly differ from the period before and after licking had been received, but was higher than the median heart rate over the whole observation period.

Effects of Different Licking Categories

For actors, but not receivers, there was a tendency for a smaller decrease in heart rate (differences between LSmeans=1.8 bpm) when they licked after solicitation compared to spontaneous licking (differences between LSmeans=2.9 bpm; Table 13.4). After licking, heart rates further decreased when licking had been solicited but remained unchanged after spontaneous licking. Compared to the median heart rate over the whole observation period actors showed higher heart rates only during licking after solicitation. This was more pronounced in lying than in standing animals.

Licking after solicitation numerically, but non significantly reduced heart rate in receivers (difference between LSmeans=1.7 bpm). However, being spontaneously licked caused a significant increase (difference between LSmeans=3.1 bpm) which was followed by a decline of a similar magnitude. Heart rate during being spontaneously licked moreover exceeded the median over the whole observation period.

13.4.2 Loose Housed Dairy Cows

In total, 278 licking bouts fulfilling the criteria mentioned above were recorded (Table 13.5).

Overall Effects of Licking

In loose housed cows, social licking had no general effect on heart rate in both actors and receivers (Table 13.6).

Effects of Different Licking Categories

In actors, no significant influence on heart rate was found for any of the licking categories investigated (Table 13.7), but average heart rates were numerically lower during than before and after performing licking which had been solicited (differences of LSmeans=1.5 and 2.0 bpm, respectively, not significant). However, the general activity significantly influenced changes in heart rate for spontaneous licking. A marked decrease and subsequent increase (difference of LSmeans=3.0 and 3.4 bpm) was observed in feeding animals, whereas lying animals showed a reverse pattern, i.e. increase during the licking period.

TABLE 13.4 Effects of licking category and activity on heart rates in actors and receivers – tied cows (LSmeans).

Licking category			Heart rate (bpm)											p–value licking (l)	p–value activity (a)	p–value l*a
			Overall			Feeding		Standing		Lying		Lying/feeding				
			during licking	no licking	Δ bpm	during licking	no licking	during licking	no licking	during licking	no licking	during licking	no licking			
Actor	after solicitation	Licking vs. 5 min before	84.4	86.2	–1.8	–	–	78.5	78.0	9.3	94.4	–	–	.252	.041	.163
		Licking vs. 5 min after	**84.4**	**81.1**	**+3.3**	–	–	78.6	76.4	9.2	84.8	–	–	**.000**	.011	.071
		Licking vs. Ref. period	**83.9**	**78.4**	**+5.5**	–	–	78.1	75.8	89.7	81.0	–	–	**.000**	.165	*.013*
	spontaneous	Licking vs. 5 min before	**78.5**	**81.4**	**–2.9**	–	–	78.5	81.4	–	–	–	–	**.000**	–	–
		Licking vs. 5 min after	78.8	79.9	–1.1	–	–	78.8	79.9	–	–	–	–	.182	–	–
		Licking vs. Ref. period	78.4	78.5	–.1	–	–	78.4	78.5	–	–	–	–	.960	–	–
Receiver	after solicitation	Licking vs. 5 min before	76.2	77.9	–1.7	76.7	78.5	75.7	77.4	–	–	–	–	.140	.458	.967
		Licking vs. 5 min after	76.1	75.8	+.3	77.0	78.0	75.0	73.6	–	–	–	–	.729	.008	.091
		Licking vs. Ref. period	76.8	76.9	–.1	78.8	79.1	74.7	74.7	–	–	–	–	.887	.025	.837
	spontaneous	Licking vs. 5 min before	**78.7**	**75.6**	**+3.1**	82.4	82.5	75.1	77.0	75.0	74.7	69.8	8.5	**.033**	.002	.107
		Licking vs. 5 min after	**77.7**	**75.3**	**+2.4**	82.8	82.7	75.4	73.1	72.1	7.1	8.5	75.4	**.048**	.001	.484
		Licking vs. Ref. period	**78.0**	**75.0**	**+3.0**	82.9	82.3	75.0	73.0	72.9	68.5	81.2	76.4	**.014**	.001	.380

Notes: bold: significant effects of licking; italics: significant effects of interaction licking*activity; bpm: beats per minute.

TABLE 13.5 Number of valid licking bouts per category in loose housed animals.

	Licking categories	Activity	Number of licking bouts
Actor	Spontaneous	Feeding	23
		Standing	25
		Lying	22
	Solicited	Feeding	3
		Standing	18
		Lying	11
	After agonistic interaction	Feeding	3
		Standing	12
		Lying	1
	Spontaneous	Feeding	20
		Standing	13
		Lying	60
	Solicited	Feeding	5
		Standing	19
		Lying	32
	After agonistic interaction	Feeding	1
		Standing	10
		Lying	–
Total			278

TABLE 13.6 Effects of licking on heart rates in actors and receivers – loose housed cows.

		Differences between LS means (beats per minute)	F–value	Standard error	p–value
Actors	Licking vs. 5 min before	–.4	$F_{1,202}$=.258	.73	.612
	Licking vs. 5 min after	–.7	$F_{1,196}$=1.049	.71	.307
	Licking vs. Reference period	+.1	$F_{1,187}$=.021	.833	.884
Receivers	Licking vs. 5 min before	–.9	$F_{1,285}$=2.051	.59	.153
	Licking vs. 5 min after	–.7	$F_{1,287}$=1.390	.57	.239
	Licking vs. Reference period	+.7	$F_{1,272}$=1.603	.58	.207

On the other hand, heart rate in receivers was only influenced when they had been licked after solicitation. Heart rate during licking after solicitation was lower than before and after the licking bout (difference of LSmeans=2.4 and 2.7, respectively) and compared with the median heart rate over the whole observation period (of the same activities). This decrease was only observed in feeding and standing animals and did not occur in lying animals.

TABLE 13.7 Effects of licking category and activity on heart rates in actors and receivers – loose housed cows (LSmeans).

Licking category			Heart rate (bpm)									p–value licking (l)	p–value activity (a)	p–value l*a
			Overall			Feeding		Standing		Lying				
			during licking	no licking	Δ bpm	during licking	no licking	during licking	no licking	during licking	no licking			
Actor	after solicitation	Licking vs. 5 min before	89.7	91.2	–1.5	91.6	95.6	9.2	91.3	87.4	86.7	.306	.008	.489
		Licking vs. 5 min after	89.6	91.6	–2.0	91.1	95.9	89.8	9.4	88.0	88.4	.159	.053	.470
		Licking vs. Ref. Period	9.0	9.8	–.8	92.2	95.9	9.2	93.7	87.5	83.0	.490	.000	*.001*
	spontaneous	Licking vs. 5 min before	91.5	91.8	+.3	93.7	96.7	9.8	92.0	9.0	86.6	.708	.000	*.004*
		Licking vs. 5 min after	93.9	92.3	+1.6	93.9	97.3	9.7	91.4	9.1	88.2	.334	.000	*.022*
		Licking vs. Ref. Period	91.6	91.3	+.3	94.5	97.3	91.5	93.8	89.3	83.0	.556	.045	*.000*
	after agonistic interaction	Licking vs. 5 min before	92.3	91.5	+.8	93.5	96.0	9.3	92.0	9.3	86.6	.745	.174	.545
		Licking vs. 5 min after	91.8	91.8	.0	92.4	96.3	9.8	9.9	92.2	88.1	.981	.213	.415
		Licking vs. Ref. Period	92.9	9.9	+2.0	93.2	96.4	92.2	93.8	93.4	82.5	.269	.045	.022
Receiver	after solicitation	Licking vs. 5 min before	**88.3**	**9.7**	**–2.4**	92.1	95.0	86.6	91.1	86.0	86.2	**.010**	.000	*.043*
		Licking vs. 5 min after	**87.8**	**9.5**	**–2.7**	91.1	95.0	86.7	9.5	85.7	86.0	**.002**	.000	*.044*
		Licking vs. Ref. Period	**87.3**	**89.7**	**–2.4**	9.9	93.1	85.8	91.8	85.2	84.2	**.002**	.000	*.000*
	spontaneous	Licking vs. 5 min before	9.3	9.8	–.5	94.6	95.4	89.8	9.1	86.4	86.2	.455	.000	.676
		Licking vs. 5 min after	9.0	9.6	–.6	94.4	95.5	88.8	9.4	86.7	86.0	.343	.003	.567
		Licking vs. Ref. Period	9.1	89.9	+.2	94.7	93.8	89.1	91.8	86.6	84.0	.649	.000	*.001*
	after agonistic interaction	Licking vs. 5 min before	92.5	92.1	+.4	96.2	94.3	88.9	89.9	–	–	.863	.022	.538
		Licking vs. 5 min after	92.9	92.5	+.4	96.8	95.3	89.0	89.7	–	–	.852	.003	.567
		Licking vs. Ref. Period	94.1	93.3	+.8	96.5	94.0	91.8	92.5	–	–	.643	.135	.407

Notes: bold: significant effects of licking; italics: significant effects of interaction licking*activity; bpm: beats per minute.

13.5 DISCUSSION

In tethered cows, heart rates showed expected patterns in that lying animals had lower heart rates than more active animals (at least in the receivers, but numbers of lying actors were very low). Interestingly, licking more clearly led to a decrease in heart rates in actors than in receivers although the number of analysed active lickings was much lower (38 vs. 79 received lickings). Moreover, the effect was stronger when the actors performed the behaviour spontaneously instead of reacting to a request of the other cow. Possible explanations of this finding may relate to the specific conditions in tie stalls: The forced spatial vicinity between the cows and impossibility of avoidance may lead to social tension so that licking is used for appeasement (Reinhardt, 1980; Sato et al., 1991; Waiblinger et al., 2002; Emmerig, 2004). Fraser and Broom (1990) furthermore speculate that social grooming may be a way to cope with restrictive conditions by self-narcotisation, as opioids from studies with rats and primates are known to be involved in allogrooming both in the receiver and the actor (Keverne et al., 1989; Niesink and van Ree, 1989). In both cases of appeasement and self-narcotisation a decrease in heart rate were to be expected which is consistent with our results for actors. In receivers, variability in responses was higher, even including heart rate rises in cows being licked spontaneously.. From the behavioural observations that were unsystematic in this regard, it especially appeared that cows being licked spontaneously sometimes showed a mild startle response after being touched by the other cow. Both, receivers being licked spontaneously and actors licking after solicitation, had higher heart rates than during comparable periods without licking (median heart rate). This might indicate that the licking occurred during a stressful period or was induced by it. However, as the data basis is not strong enough for far-reaching interpretations this aspect deserves further investigation. Overall the calming or relaxing effect of licking could only clearly be confirmed in actors.

In loose housed cows, lying animals also had lower heart rates compared to other activities. No general calming or relaxing effect of licking was observed in both actors and receivers. Yet, looking only at animals which performed spontaneous licking during standing/feeding, a reduction of heart rate was present. This is in line with the results of the tethered cows, where all cows performing spontaneous licking were standing. On the other hand, the increase in heart rate in lying actors may be explained by the physical activity of the animal performing the behaviour (Baldock et al., 1988; Marchant et al., 1997). Interestingly, in receivers a calming effect was only observed when they had solicited licking. This effect was again restricted to standing and feeding animals. Tension-reduction after agonistic interactions as suggested by Boccia et al. (1989) could not be confirmed.

Calming or relaxing effects could be shown in loose housed cows, but were restricted to specific categories of licking and/or the underlying general activity with no or even arousing effects in lying animals. Actors might perceive licking pleasant or experience tension-reduction if they 'voluntarily' perform this behaviour while this is the case for receivers when they are being licked after solicitation.

13.6 CONCLUSIONS

We indeed found indications of calming effects of licking, although the findings from tied and loose housed animals are not completely consistent. Altogether it is very likely that licking and being licked usually increases well-being of the individual, no matter if receiver or, according to our results from tie stalls, actor. However, on herd level it is questionable whether high numbers of animals performing licking always reflect an overall better welfare state compared to animals of a herd with lower licking frequencies. In the literature it is suggested that licking may also be carried out in order to reduce social tension or to reduce individual stress responses to restrictive conditions. Our results are consistent with these hypotheses in that in the tie stalls more pronounced and consistent heart rate reductions were found in actors, especially in those that licked spontaneously. Moreover, from the tie stall data there was an indication that licking occurred during a stressful period or was induced by it, although this needs further investigation.

We conclude that social licking might indicate positive effects on (at least part of) the animals involved (in terms of reducing stress or stabilising social bonds) but on the basis of changes in heart rate at present cannot be regarded as a true indicator of positive emotions. Therefore, the use of social licking as a welfare measure in on-farm assessment protocols is not recommended.

14

RELIABILITY OF ON-FARM BEHAVIOURAL OBSERVATIONS OF ABNORMAL BEHAVIOUR IN VEAL CALVES

E.A.M. Bokkers, H. Leruste, L.F.M. Heutinck, M. Wolthuis-Fillerup,
J.T.N. van der Werf, B.J. Lensink and C.G. van Reenen

14.1 SUMMARY

The objective was to study inter-observer and test-retest reliability of different behavioural observations to be used in an on-farm animal welfare monitoring system for veal calves. Twenty-three veal calf farms varying in size, housing system, feeding regime and age of the calves were visited twice with two observers simultaneously. Behaviour was recorded 20 min before and 20 min after feeding in eight pens per farm.

For most behavioural elements recorded around feeding farms differed significantly and inter-observer and test-retest reliabilities were high and significant as well.

The behavioural observations around feeding were feasible, distinctive and reliable to perform on-farm. These methods are promising tools to use in an animal welfare monitor for veal calves.

14.2 INTRODUCTION

Abnormal behaviour is widely accepted as an indicator for poor welfare (Fraser and Broom, 1997; Anonymous, 2001) while play behaviour is an indicator for good welfare (Fagen, 1981; Newberry et al., 1988). Play behaviour of calves has been studied extensively (e.g. Jensen et al., 1998; Jensen and Kyhn, 2000) as well as abnormal behaviour. Veal calves typically develop abnormal oral behaviour, comprising the following four behavioural elements: tongue playing, tongue rolling, sham ruminating, and persistent biting/sucking on substrates such as bars and troughs (Bokkers and Koene, 2001). Although housing system may affect the frequency of abnormal oral behaviour in veal calves (Bokkers and Koene, 2001), it has been clearly demonstrated that abnormal oral behaviour to a large extent evolves due to a lack of appropriate roughage in the diet (Heeres et al., 2000; van Vuuren et al., 2004). Other abnormal (oral) behaviours in veal calves include cross-sucking and excessive self-licking. Cross sucking, defined as one calf sucking the ear, mouth, scrotum, prepuce, tail, udder area or navel of another calf (Lidfors, 1993), is seen most often in young calves that are separated from their mother. Persistent preputial sucking may detrimentally affect the prepuce (swelling, irritation, inflammation) of the calf being sucked and the calves that suck may risk a poor health and reduced growth due to drinking urine (de Wilt, 1985). Self licking is a normal behaviour for a calf, but it can develop to an abnormal, excessive level, especially when a calf is kept in social isolation (Terosky et al., 1997; Bokkers and Koene, 2001).

Until now, relatively simple and feasible behavioural measures in veal calves have been sufficiently validated under experimental conditions. Before behavioural observation methods, however, can be considered for inclusion into an animal welfare monitoring system, they have to be studied for feasibility and reliability under commercial conditions. The aim of this study was to investigate the feasibility of behavioural observation methods to study spontaneous behaviour in veal calves kept under commercial conditions, and to analyse inter-observer reliability and test-retest reliability for different variables. Variables that appear feasible and reliable may be suitable for an on-farm animal welfare monitoring system.

14.3 METHODS

Twelve veal farms (11 farms with small groups, i.e. < 20 calves per pen, and one farm with large groups and automatic milk dispensers) were included in this study. These farms varied according to type and origin of calves, group size, size of the farm, diet (amount of milk replacer and amount and type of solid feed), climate control, day light intensity, and management. They were assumed to represent a cross-section of veal farms in the Netherlands.

14.3.1 Farm Visits

Each farm was visited twice to collect data, with a 1–3 days interval between two visits to be able to study test-retest reliability of the observations. Farms were visited two weeks prior to slaughter. During all visits, data were collected by two observers simultaneously in order to study inter-observer reliability. Four observers (two men and two women) visited the farms in different combinations, but always the same combination of observers within a farm. They were wearing workwear of the farmer. Although the observers were experienced in behavioural research, they completed a training assessment with videos and photos of calf behaviour and practised together at a farm beforehand. One woman did observations at all farms.

14.3.2 Observations of Spontaneous Behaviour

Behaviour in the homepen was observed in eight randomly selected pens 20 min before and 20 min after evening feeding. By performing the behavioural observations relative to a fixed event, i.e. feeding, results between farms and days could be compared. The ethogram included two postures (i.e., standing and lying) and 20 behavioural activities. In case of small groups (< 20 calves per pen) four pens were observed at the same time. The observers stood next to each other in the feeding corridor. Before starting the observation, a 5 min adaptation period was maintained. The observers had 30 s of observation time per pen and they switched to the next pen at the same moment (clockwise). In total 10 scans per pen per observation were recorded. After 20 min the observers moved to the next four pens to repeat this procedure. After the calves had been fed, behavioural observations started at the first location again. At farms with large groups, two pens were observed (2 × 20 min for each pen around the same time as for small groups).

14.3.3 Data Analysis

Data were analysed at pen level with the statistical software package Genstat (2005). Spearman's rank correlations were calculated as a measure for reliability between days and between observers. Correlation coefficients were considered low when below 0.4, moderate when 0.4 to 0.7, high when 0.7 to 0.9, and very high when 0.9 and higher (Martin and Bateson, 1993). Inter-observer reliabilities were analysed for the three pairs of observers. Overall, inter-observer reliabilities were analysed with Kendall's coefficient of concordance. Test-retest reliability was analysed at pen level with the mean over the two observers per pen as unit, and at farm level with the mean over observers and pens per farm as unit.

Except for lying idle, standing idle and walking, behaviour is expressed without distinguishing between lying and standing. Play behaviour is the sum of running, jumping, mounting, and butting behaviour. Comfort behaviour is the sum of stretching, scratching, nose licking and self-licking behaviour. The level of each behaviour was calculated for the 20 min period before and the 20 min after feeding. Farm effects for all variables were analysed with the Kruskal-Wallis test.

14.4 RESULTS

Significant farm effect were found for a number of behaviours, including: lying idle ($p<0.001$ before and after feeding on days 1 and 2), tongue rolling/tongue playing (before feeding: $p<0.001$ on day 1 and $p<0.05$ on day 2; after feeding: ns), manipulating fence (before feeding: $p<0.01$ on days 1 and 2; after feeding: ns on day 1 and $p<0.05$ on day 2), manipulating feeder ($p<0.01$ before and after feeding on days 1 and 2), manipulating pen mate (before feeding: $p<0.001$ on days 1 and 2; after feeding: $p<0.001$ on day 1 and ns on day 2), and (sham) rumination ($p<0.01$ before and after feeding on days 1 and 2). For play behaviour, no significant farm effect was found.

Test-retest correlations differed per behaviour. Before feeding, significant test-retest correlations were obtained for lying idle (days within pen: $r=0.63$; $p<0.001$; $n=90$; days within farm: $r=0.34$; ns; $n=12$), tongue rolling/tongue playing (days within pen: $r=0.41$; $p<0.001$; $n=90$), manipulating fence (days within pen: $r=0.56$; $p<0.001$; $n=90$; days within farm: $r=0.92$; $p<0.001$; $n=12$), manipulating feeder (days within pen: $r=0.55$; $p<0.001$; $n=90$; days within farm: $r=0.73$; $p<0.01$; $n=12$), manipulating pen mate (days within pen: $r=0.43$; $p<0.001$; $n=90$; days within farm: $r=0.64$; $p<0.05$; $n=12$), and (sham) rumination (days within pen: $r=0.33$; $p<0.01$; $n=90$; days within farm: $r=0.60$; $p<0.05$; $n=12$). After feeding, test-retest reliabilities for tongue playing/tongue rolling were non-significant. After feeding significant test-retest reliabilities were found for lying idle (days within pen: $r=0.36$; $p<0.01$; $n=90$; days within farm: $r=0.67$; $p<0.05$; $n=12$), manipulating fence (days within pen: $r=0.26$; $p<0.05$; $n=90$; days within farm: $r=-0.02$; ns; $n=12$), manipulating feeder (days within pen: $r=0.66$; $p<0.001$; $n=90$; days within farm: $r=0.79$; $p<0.01$; $n=12$), manipulating pen mate (days within pen: $r=0.37$; $p<0.001$; $n=90$; days within farm: $r=0.36$; ns; $n=12$), and (sham) rumination (days within pen: $r=0.39$; $p<0.001$; $n=90$; days within farm: $r=0.75$; $p<0.01$; $n=12$).

Inter-observer correlations (Kendall's coefficients of concordance, $n=64$) were generally high and significant for the different behaviours before and after feeding: lying idle (before and after feeding: 0.73 and 0.73, respectively; $p<0.01$), tongue playing/tongue rolling (before and after feeding: 0.75 and 0.73, respectively; $p<0.01$), manipulating fence (before and after feeding: 0.73 and 0.71, respectively; $p<0.01$), manipulating feeder (before and after feeding: 0.76 and 0.74, respectively; $p<0.001$), manipulating pen mate (before and after feeding: 0.71 and 0.71, respectively; $p<0.01$), play behaviour (before and after feeding: 0.66 and 0.56, respectively; $p<0.01$), and (sham) rumination (before and after feeding: 0.71 and 0.66, respectively; $p<0.01$).

14.5 CONCLUSIONS

Most inter-observer and test-retest reliabilities for the behavioural measures were significant, and correlation coefficients were higher at farm level than at pen level. This is relevant for the development of an animal welfare monitoring system aimed at estimating animal based parameters at farm level. Behavioural observations around feeding were feasible to perform on farm, although there were some minor practical constraints. It can be concluded that the observation method for spontaneous behaviours represents a reliable tool for utilization in an animal welfare monitoring system.

15

MEASUREMENT OF HUMAN-ANIMAL RELATIONS IN VEAL CALVES HOUSED IN LARGE GROUPS

H. Leruste, J. Lensink and C.G. van Reenen

15.1 SUMMARY

The human-animal relationship is an important component of the welfare of farm animals. Under experimental conditions, relatively simple and feasible behavioural measures of the human-calves relationship have been validated. This study aims at determining simple and valid tests which would be good candidates to integrate in an animal-based-on-farm assessment tool. They should reflect the actual level of fear of human of the group of calves, be repeatable at short and medium term (test-retest), work for large groups (between 20 and 60 calves) and not be affected by an observer effect.

Observations were performed on 20 farms in France. Calves were housed in large groups and fed with an automatic milk dispenser. Observations were performed on 4 days: 2 observations at a two day interval at week 9 and 2 observations at a two day interval at week 13. Observations were performed simultaneously by two trained observers. Four categories of measures were performed: reaction of the group to humans standing in front of the pen (% of calves reacting), reaction of the group towards unfamiliar persons entering the pen (% of calves escaping and following), voluntary approach of the group of motionless unfamiliar persons (latency for the 1st touch, mean number of calves close to the persons and speed-of-approach of the group) and reaction of the calves to being touched by a human (% of calves that could be touched).

Observers agreed for all the measures evaluated: reaction to humans in front of the pen ($r=0.96$; $p<0.001$; $n=20$), % of calves escaping ($r=0.84$; $p<0.001$; $n=20$), latency for the 1st touch ($r=0.99$; $p<0.001$; $n=20$), mean number of calves close to the observers ($r=0.98$; $p<0.001$; $n=20$), speed-of-approach score ($r=0.96$; $p<0.001$; $n=20$) and number of calves touched ($r=0.93$; $p<0.001$; $n=20$). Inter-observer reliability was a bit lower for the percentage of calves following the observers ($r=0.70$; $p<0.001$; $n=20$). The mean number of calves close to the observers and the speed-of-approach score showed good repeatability at short term (respectively $r=0.74$; $p<0.001$; $n=61$ and $r=0.66$; $p<0.001$; $n=61$) and at

medium term (respectively $r=0.63$; $p<0.001$; $n=66$ and $r=0.54$; $p<0.001$; $n=66$). The other measures showed medium to low consistencies at short and medium term. We observed a habituation of the calves to the tests with an increase of the approach between day 1 and day 4.

The latency to touch the observers, the mean number of calves close to the motionless observers and the percentage of calves escaping and following the observers when they are crossing the pen are good candidates for the welfare monitoring system. They permit to assess the voluntary approach of the group and the reaction to humans moving. The other measures don't seem to be reliable. This experiment permitted to propose two valid observations for on-farm assessment of human-animal relations.

15.2 INTRODUCTION

The human-animal relationship is an important component of the welfare of farm animals. Several studies show that the behaviour of the farmer has an impact on both the welfare of calves and the productivity of the veal unit. Calves learn to discriminate between people by using their previous experience with these persons. De Passillé et al. (1996) found that calves, tested in their home pen, showed longer latency to touch a person who had previously handled them negatively. The daily behaviour of the farmer has an impact on the emotional response of calves to handling and transport (Lensink et al., 2001b). Calves originated from veal units where the farmer was considered as providing positive behaviours needed fewer efforts to be loaded/unloaded in a truck, had lower heart rates during loading/unloading and less incidents at slaughterhouse. Their carcasses were also fund to be paler and their meat had a lower pH. Lensink et al. (2000a) found that farmers who had positive contacts with their calves had higher productivity levels in terms of daily weight gain and feed conversion.

Under experimental conditions, validity of simple behavioural measures reflecting responsiveness of veal calves to humans has been examined in different ways. Calves housed individually or in group pens of two calves were subjected to an experimental regime putatively affecting fear of humans, and based on either a positive or a neutral handling treatment by the stockpersons (Lensink et al., 2000a, 2001b). The calves' fear reaction to humans was measured by simple approach and avoidance tests during and shortly after milk feeding in the home environment. These responses were found to be significantly influenced by the treatment, providing evidence for their predictive validity. Most importantly, the effect of the handling treatment did not significantly interact with the effect of housing (Lensink et al., 2001a), suggesting that findings obtained in individually housed calves may also apply to group-housed ones. Secondly, individual differences in the behavioural reactivity of veal calves to simple behavioural tests in the home pen have been shown to be associated with individual differences in their responsiveness to a widely used standard test measuring voluntary approach behaviour to a human in a test arena (Lensink et al., 2000a, 2003). Since the behavioural response of farm animals to this latter test is generally believed to reflect fear of humans (Hemsworth, 2003), this finding provides evidence for concurrent validity of simple measures recorded in the home environment. Additional support for the validity of these measures, as well as for their feasibility when used on commercial veal farms, was obtained in a study demonstrating a clear link between the stockperson's behaviour towards the calves during feeding (in terms of putatively neutral, positive or negative interactions) and fear responses of calves to humans as recorded in approach and avoidance tests (Lensink et al., 2000b).

Therefore, based on current knowledge, we assume that relatively simple and feasible behavioural measures in veal calves in the area of the human-animal relationship have been sufficiently validated. Human-animal relations can be easily assessed on experimental conditions. Only few studies have been performed so far on feasibility and reliability of these measures under truly commercial conditions. The development of welfare monitoring schemes (EU project) creates a need for a valid method to assess human-animal relations on farm.

Lensink et al. (2003) studied the repeatability (test-retest reliability) of behavioural responses to a human of group-housed veal calves under semi-commercial conditions. They measured the reaction of calves when touched during feeding and found that it was repeatable in time. Rousing et al. (2005) also tested the repeatability of the reaction to human approach of group-housed calves during feeding. They found that satisfying inter-observers agreement and between days consistency. These tests seem then valid and applicable on farm. Nevertheless, in these studies the group size examined was either 2 or 5 (in Lensink et al., 2003) or 3 to 10 (in Rousing et al., 2005). A group size between 5 and 7 is currently the most common in the European Union, however, husbandry systems with much larger groups (between 20 and 60 calves) fed by an Automatic Milk Dispenser (AMD) are gaining popularity. Therefore, feasibility of the measurement in this system must be studied.

This study aims at determining simple and valid tests which would be good candidates to integrate in an animal-based-on-farm assessment tool. They should reflect the actual level of fear of human of the group of calves, be repeatable at short and medium term (test-retest) and shouldn't be affected by an observer effect.

15.3 METHODS

As presented above, four categories of measures were performed:

- Reaction of the group to humans standing in front of the pen

- Reaction of the group toward an unfamiliar person entering the pen
- Voluntary approach of the group toward an unfamiliar person
- Reaction of the calves to being touched by a human

On farm observations were performed on 20 farms in France between October 2005 and March 2006. Calves were housed in large groups (between 20 and 64 calves per pen) and fed with an automatic milk dispenser (AMD). Observations were performed on 4 days: 2 observations at a two day interval at week 9 (9 weeks after the arrival of the calves in the farm) and 2 observations at a two day interval at week 13. Between 1 and 6 pens were observed per farm (for a total of 67 pens observed). Observations were performed simultaneously by two trained observers.

The two observers calmly entered the building and staid for 15 seconds in front of the pen. They then entered the pen and crossed it in a standardized manner, they stood motionless for 3 minutes at the back of the pen and let the calves approach. At the end of the 3 minutes they tried to touch the forehead of the calves which were in an arm's length from them. These four tests permitted to build 7 measures: % of calves reacting to humans standing in front of the pen, % of calves escaping the observers crossing the pen, % of calves following the observers crossing the pen, latency for the first calf to touch the observers, mean number of calves standing close to the motionless observers (during the 3 minutes test), speed-of-approach score of the group, and % of calves in an arm's length distance from the observers that could be touched.

15.3.1 Data Analysis

Inter-observer repeatability was assessed through spearman rank correlations. Between-day consistency was assessed at short term (day1–day2 and day3–day4) and at medium term (week9–week13) using spearman rank correlations.

Analyses of variance were used to assess the effects of environmental parameters (size of the group, feeding system…) on the observation and to analyse farm and pen effects. The model included:

- fixed variables: day of observation (or week of observation) and number of calves in the pen;
- random variable: farm;
- interactions: farm and day (or week), number of calves and day (or week).

The evolution of the variables in time was assessed thanks to a repeated measurement analysis.

15.4 RESULTS

A farm effect was found for all measures except for the number of calves touched.

Observers agreed for all the measures evaluated: reaction to humans in front of the pen (r=0.96; p<0.001; n=20), % of calves escaping (r=0.84; p<0.001; n=20), latency for the 1st touch (r=0.99; p<0.001; n=20), mean number of calves close to the observers (r=0.98; p<0.001; n=20), speed-of-approach score (r=0.96; p<0.001; n=20) and number of calves touched (r=0.93; p<0.001; n=20). Inter-observer reliability was a bit lower for the % of calves following the observers (r=0.70; p<0.001; n=20).

The mean number of calves close to the observers and the speed-of-approach score showed good repeatability at short term (respectively r=0.74; p<0.001; n=61 and r=0.66; p<0.001; n=61) and at medium term (respectively r=0.63; p<0.001; n=66 and r=0.54; p<0.001; n=66). The other measures showed medium to low consistencies at short and medium term: reaction to humans standing in front of the pen (r=0.44; p<0.001; n=56 and r=0.54; p<0.001; n=60), % of calves escaping (r=0.43; p<0.001; n=59 and r=0.36; p=0.003; n=66), % of calves following (r=0.33; p=0.01; n=60 and r=0.60; p<0.001; n=65), latency to touch the observers (r=0.47; p<0.001; n=61 and r=0.56; p<0.001; n=66) and % of calves touched (r=0.25; p=0.06; n=58 and r=0.29; p=0.02; n=64). We observed a habituation of the calves to the tests, they were following more the observers and approaching faster and were more numerous to stand close to the observer at week 13 than at week 9 (p<0.001).

A pen effect was found for all measures: between 12% and 24% of farms showed a pen effect. The breed of the calves did affect all measures except the mean number of calves close to the observers and the speed-of-approach score. For all measured were a difference was found, Crossbred calves seemed less fearful of humans than Holstein calves. The size of the group influenced the proportion of calves escaping (p=0.07), the mean number of calves close to the observers (p<0.001) and the speed-of-approach score (p<0.001).

Most of the measures are linked. Groups which are reacting strongly to the presence of humans are reacting also when they are crossing the pen (escaping: r=0.20; p<0.001 and following: r=0.33; p<0.001) they approach faster of the observers (lower latency to touch: r=-0.24; p<0.001 and higher speed-of approach score: r=0.16; p=0.01) and are more numerous around the observers (means number of calves: r=0.20; p=0.002). The mean number of calves touched is not linked with all the other variables but this might be due to a low variability within this measure. Speed-of approach score is highly correlated to the mean number of calves close to the observers (r=0.95; p<0.001).

15.5 CONCLUSIONS

The latency to touch the observers combined with the mean number of calves close to the observers in the three minutes test are the best candidates regarding the different criteria. They permit to assess the voluntary approach of the group. They could be completed by an assessment of the reaction to humans moving thanks to the percentage of calves escaping and following the observers when they are crossing the pen. The reactions to the human presence in the building and to humans trying to touch the calves don't seem to be reliable measures.

This experiment permitted to develop two valid observations (providing four valid measures) for on-farm assessment of human-animal relations. Final decisions on good candidates will be made after the analysis of the data on the assessment of human-animal relations in small groups.

16

ASSESSMENT OF HUMAN–ANIMAL RELATIONSHIPS IN DAIRY COWS

I. Windschnurer, C. Schmied, X. Boivin and S. Waiblinger

16.1 SUMMARY

The animal's relationship to humans (AHR), i.e., the perception of humans and of the interactions with humans, was shown to have a major impact on animal health, productivity, and welfare and thus is an important parameter to include in on-farm welfare assessment (Hemsworth and Coleman, 1998; Boivin et al., 2003; Waiblinger et al., 2003; Winckler et al., 2003). The aim of this project was to test different tests for measuring the Animal Human Relationship in dairy cows with respect to their reliability and partly validity.

In loose-housed dairy cows, besides already existing measures some simple tests were developed and both were evaluated with respect to validity and reliability (inter-observer reliability, between-experimenter, and test-retest repeatability). This was done in the course of two separate studies. One study on 33 commercial dairy farms allowed evaluating the validity of the tests for measuring AHR by analysing correlations between human handling behaviours during milking and the animals' responses during the tests. In the course of the other study, performed on 16 commercial dairy farms, the reliability of the tests was investigated as well as inter-test relationships.

Six behavioural tests were evaluated: avoidance distance in the barn (AD), avoidance distance at the feeding place (ADF), approach to a stationary human in the barn (APP), reactions of lying animals to a human passing or touching them (LP; LT) and reactions of animals standing in an alley to a human walking through the herd and trying to touch them (WT).

Regarding validity, AD showed the strongest and highest number of correlations with milker behaviour. The more positive handling behaviours where recorded for milkers, the lower was the average AD, the median value of AD and the proportion of animals with a higher AD (>1m, >2m) ($r = -0.42$ to -0.49; $p<0.05$ to 0.01; N= 33). Some ADF measures also showed significant negative correlations with positive milker behaviour (average ADF: $r = -0.38$; % of animals with ADF>0.5m: $r = -0.35$; both $p<0.05$, N= 33). The other

tests showed no significant correlation to milker behaviour, although a tendency was found for WT. With regard to inter-test relationships, also a measure of convergent validity, high correlations (partly > 0.8) were found between AD, ADF, and WT.

With regard to between-experimenter repeatability (level of consensus between results when the test is performed twice by different experimenters) also AD and ADF turned out to be the most reliable measures. At farm level, both avoidance distance measures (AD and ADF) showed mainly high (r≥0.8) to moderate (r≥0.7<0.8) repeatability. In contrast, the other tests did not show such a high repeatability (no to moderate repeatability, depending on the measure).

Inter-observer reliability (level of consensus when two persons observe and record the same event/test at the same time) was high for all of the tests (e.g., for AD r= 0.99 / 0.99; ADF: r= 0.97 / 0.98).

With regard to tie-stall systems, only one farm could be visited. Three experimenters subjected 54 animals to ADF. The inter-observer reliability was comparable to ADF in loose housing, but the repeatability was lower. However, at an individual level the repeatability was also lower in loose-housed animals (= consistency in individuals reactions). Thus, at farm level a sufficient repeatability can be expected.

To conclude, reliable and valid measures have been found for the assessment of the animal-human relationship in dairy cows in different housing systems.

16.2 INTRODUCTION

The animal's relationship to humans, i.e., the perception of humans and of the interactions with humans, was shown to have a major impact on animal health, productivity, and welfare and thus is an important parameter to include in on-farm welfare assessment (Hemsworth and Coleman, 1998; Boivin et al., 2003; Waiblinger et al., 2003; Winckler et al., 2003; Waiblinger et al., 2006).

Measuring animals' reactions to humans enables us to reach conclusions about how they perceive specific human beings or people in general. The animal's reactions reflect a mixture of different emotions belonging to the two main dimensions pleasant (e.g., security) and unpleasant (e.g., fear, pain) (Waiblinger et al., 2006). In different studies, many tests have been used for measuring the relationship of cattle to humans. However, studies using tests potentially feasible for an on-farm welfare assessment scheme are rare (for loose-housed dairy cows: e.g., De Rosa et al., 2003; Waiblinger et al., 2003; Rousing and Waiblinger, 2004). Even less is known about the repeatability of the animals' reactions to different unknown humans. Furthermore, some of the tests used so far are still relatively time-consuming.

Therefore, the aims of the project was to (1) develop new, simple, and quick tests for dairy cows, and (2) to study the reliability and partly validity of the new tests and some of the established tests. This chapter will emphasise on the reliability study, but will also give excerpts, e.g., some results, from analyses of validity.

Reliability comprises different aspects. Intra-observer reliability (degree of consensus when one assessor measures the same event at the same time), inter-observer reliability (degree of consensus when more assessors measure the same event at the same time) and test-retest reliability (=repeatability; consistency of the measurement when it is repeated within a brief time period). Because of the nature of our measures, where the assessor is at the same time the test person to whom reactions of animals are measured, the consistency in animals' reactions to the same human (part of the variability in test-retest reliability) and especially to different humans (e.g., assessors in on-farm assessment) is highly relevant. We tested this by repeating tests with two different test persons and refer to this aspect of reliability as between-experimenter repeatability (the experimenter being test person and observer at the same time).

For testing intra- and inter-observer reliability the assessor(s) have to observe the same event at the same time. With regard to inter-observer reliability, this was performed by one person (the experimenter) testing the animals and noting the reaction and a second person (or also a third person) observing this test from a distance and noting the animals' reactions. Intra-observer reliability cannot be tested live. However, the assessment of reactions from video recordings in order to evaluate intra-observer reliability was not possible for most of the tests. Yet, a high inter-observer reliability indicates that it is possible to reach high intra-observer reliability too. Thus, we concentrated on the first.

Another important aspect on-farm welfare assessment has to deal with is the (long-term) stability of the parameters. The stability of a parameter is tested with longer intervals and says something about potential real changes. The stability could be low, despite high test-retest reliability. This aspect was not part of our project, but a recent study has shown a good stability over a 5-month period in some tests assessing the animal-human relationship (avoidance distances) in dairy cows (Winckler et al., 2007).

16.2.1 Development and Choice of Tests

For loose-housed dairy cows, avoidance distances or avoidance reactions measured in the barn were shown to be valid and showed promising reliability (Waiblinger et al., 2002; De Rosa et al., 2003; Waiblinger et al., 2003; Rousing and Waiblinger, 2004). Although there were hints that a sufficiently high reliability may need a relatively large sample size (Waiblinger and Menke, 2003), which would not be feasible for on-farm welfare assessment, this was not investigated enough until now. Therefore, we included this measure in the battery of tests to be evaluated. Moreover, we included an avoidance distance test performed at the feeding place and an approach test to a stationary unfamiliar human (Waiblinger et al., 2003; Rousing and Waiblinger 2004). Additionally, two simpler tests were developed. These tests were studied (1) with respect to convergent validity in an on-farm survey with 33 farms by investigating associations with milker behaviour (Schmied and Waiblinger, 2006a; Waiblinger et al., 2007a). In order to assess convergent validity one inquires if 'conceptually related measures are associated with one another' (Waiblinger et al., 2006). In another on-farm study with 16 farms (2) reliability aspects and additionally inter-test relationships, i.e., also convergent validity, were investigated (Windschnurer et al., 2008). For tie-stall systems, we investigated the avoidance distance at the feeding place with regard to reliability.

16.3 METHODS

16.3.1 Farms, Animals, and Housing

For testing the reliability of tests for loose housing systems, 16 commercial dairy farms with 19 to 78 lactating cows (36.25 ± 14.46 per herd, in total 580 cows) were visited (Windschnurer et al. 2008). Depending on the tests that were performed, different numbers of animals out of the herd were tested. On one farm, cows were housed in a deep litter system and on the other farms in cubicle loose housing. The running area was made of slatted concrete floors on 11 farms and of solid concrete on 5 farms. The farms kept predominantly Simmental cows, with a few Holstein Friesian cows on some of the farms. One farm kept Holstein Friesian cows only.

For testing the reliability of tests for tie-stall systems, one farm, the Research Estate of the University of Veterinary Medicine, Vienna was chosen and 54 dairy cows of different breeds (Simmental, Brown Swiss, and Holstein) were tested.

Additionally, convergent validity was evaluated as part of a separate study with 33 commercial dairy farms with 36 ± 15 cows per herd (Schmied and Waiblinger, 2006a; Waiblinger et al., 2007a), 16 of which were the test farms in the reliability study.

16.3.2 Description of Test Procedures and Measures Calculated

The following 6 tests for assessing the animal-human relationship were evaluated.

Avoidance Distance in the Barn (AD Test)

Individual free-standing animals were approached from the front, either in the barn (in the running area or, in deep litter, in the lying area) or in the outside run, but never on pasture. The optimal starting position was in front of the animal in a distance of 3 meters. The experimenter approached the cow with a pace of 1 step per second with the arm held in an angle of about 45° in front of the body until the animal withdrew or until touching. At the moment of withdrawal, the distance between cow and the experimenter' hand (hand to muzzle / nose) was estimated in steps of 10 cm (from 300 cm to 0 cm). If an animal withdrew immediately at the moment of touching the nose or muzzle, an avoidance distance of 0.05m was assigned. If the cow remained standing still when touched, the experimenter ran her hand to the animal's cheek and tried to stroke the cheek for at least one second. If the cow allowed to be stroked, the avoidance distance was 0.00m. For farm values, the average avoidance distance (AD), the median value of AD, and the percentages of animals possible to touch, with an avoidance distance > 0.5m, > 1.0m, > 1.5m, > 2.0m, and > 3.0m were calculated.

Avoidance Distance at the Feeding Place (ADF Test)

This test is comparable to the AD test. It was performed during the main feeding times, starting 5 to 10 minutes after the animals started feeding. The test person approached individual animals standing at the feeding place (if possible, restrained in the feed barrier). After choosing an animal, it was approached from the front, one arm held in an angle of 45° in front of the body, with a speed of 1 step per second, starting from a distance of 2m (Figure 16.1). When the animal reacted by showing clear signs of avoidance or withdrawal, the distance between the test person's hand and the muzzle / nose was estimated in steps of 10 cm. If the cow allowed to be stroked, the avoidance distance was 0.00m. The calculated farm values were: the average avoidance distance (ADF), the median value of ADF, the percentage of animals possible to touch, and percentages of animals with an

FIGURE 16.1 ADF test: the test person approaches one selected animal from the front, arm held in an angle of 45° in front of the body, with a speed of 1 step per second, starting from a distance of 2m (left), until the cow withdraws or can be touched (right).

avoidance distance > 0.3m, > 0.5m, and > 1.0m. The test procedure was the same for tied dairy cows.

Approach Test (APP Test)

A central place was chosen for the test person as a test position (during our studies always with the back to the feed barrier). The number of cows standing in the barn or in a defined area around the predetermined test position (5 m, 10 m) was counted. The test person entered the barn, walked slowly to the test position, and remained standing still for 15 min. During the test duration of 15 minutes, all animals approaching to 2.5 m and all animals making contact to the test person were noted as well as latencies to 2.5m and to contact. Percentages of animals approaching to contact and average latencies were calculated.

Lypass (LP Test)

With a pace of 1 step per second, the experimenter walked along a row of cubicles with lying animals at a close distance to the end of the cubicles (20 cm). The animals' reactions to the test person were noted: Standing up yes / no at the moment when the test person approached, passed by, or a few seconds after the test person had passed behind the animal. If an animal stood up as a reaction to another animal standing up, it was excluded. Obviously sleeping animals were excluded as well. In the deep litter system the experimenter passed behind the individual cows at a distance of 20 cm.

Lame animals were noted during the farm visits and severely lame animals were excluded. Percentages of tested animals standing up relative to the tested number were calculated.

Lytouch (LT Test)

This test was performed after the LP test. Animals still lying after passing behind them were subjected to this test. The test person walked again in close distance to the cubicles with 1 step per second. Behind every lying animal the experimenter stopped in a position where the cow could see the test person and performed the following behaviours until the cow eventually stood up: (1) placing the hand on the hind quarter of the animal for 3 seconds, (2) knocking (using 4 fingers) shortly 3 times on the animals' hind quarter and (3) knocking again 3 seconds later. It was noted if animals stood up or continued lying. The percentage of tested animals standing up relative to the tested number was calculated.

Later, LP and LT were combined to one test calculating the percentage of animals standing up during both tests relative to the number of animals tested in LP (Windschnurer et al., 2008).

Walk through and Touch (WT Test)

First a long corridor of at least 15 m length was chosen in the barn. Before starting the test there should have been at least 5 free-standing animals in the alley (without e.g., feeding at the feed barrier, or standing partly in a cubicle). The free-standing animals in the alley were counted before the experimenter started to walk towards them with a speed of 1 step per second. The cows were not directly approached, but the experimenter tried to get at shoulder level with the closest cow, staying at a distance of about one arm length. Then the test person slowly lifted the arm and tried to touch the cow's shoulder for at least 2 sec. The following reactions were noted: clear avoidance reaction of the animal when the experimenter was in a distance of: >3m, <3m, <2m, <1m, withdrawal when the arm was lifted, or the animal could be touched. Percentages of animals within these categories were calculated.

16.3.3 Design of the Reliability Study

A brief description of the validity study with 33 farms can be found in Schmied and Waiblinger (2006a) and Waiblinger et al. (2007a). Below, we will concentrate on the design of the reliability study.

Loose Housing Systems

Altogether 6 different tests for assessing animal-human relationship were evaluated for reliability. They were always carried out by the same two trained experimenters during one day visits. The two experimenters (female, 1.68m, and 1.75m, respectively) always wore green overalls. Between-experimenter repeatability and inter-observer reliability were assessed on all of the farms. For investigating between-experimenter repeatability the tests were performed twice, with the order of experimenters balanced over farms, but remaining consistent within one farm for all of the tests. That is, one experimenter performed all tests first on 8 farms, while the other performed all tests first on the other 8 farms. In order to investigate the inter-observer reliability, the experimenter currently not testing the animals observed and recorded test outcomes from a distance, while the other was testing the animals.

Table 16.1: The order for studying the reliability of the 6 tests to assess the animal-human relationship (AHR) and of the 2 tests to assess general fearfulness (GF) in 16 loose-housed dairy herds.

1. AHR Test		1. GF test		2. AHR Test		3. AHR Test		4. AHR Test		5. AHR Test		6. AHR test		2. GF test	
ADF1	ADF2	Vigi1	Vigi2	App1	App2	LP1	LP2	LT1	LT2	WT1	WT2	AD1	AD2	NO1	NO2

Notes: for abbreviations of AHR tests see *Description of Test Procedures and Measures Calculated*; for abbreviations of GF tests, see above.

The six tests were performed in a fixed order on each farm (Table 16.1). In addition to these tests, two tests to potentially assess general fearfulness (GF) were also performed. First, for assessing the level of vigilance two groups of animals feeding at the feeding place were filmed for approximately 6 minutes each, while the experimenters left the barn (Vigi1 and Vigi2). The second test was a novel object (NO) test that was carried out twice. For more information about these tests see Chapter 22.

Tie-stall Systems

Three experimenters (female, 1.68m, 1.68m, and 1.75m, respectively) performed the study. Each of the 54 tested cows was tested three times (ADF1 to 3), once by each experimenter, alternatively, in a balanced order, while the two other test persons were observing and estimating the distance as well. Other tests performed during the one-day farm visit were LP and LT. However, these tests will not be further discussed for tethered cows, as there was no variability with only 2 cows standing up during LT in the first test session. The order of tests was as follows:

LP1; LT1; ADF1; LP2; LT2; LP3; LT3; ADF2; ADF3

Due to time constraints and because a lot of animals were lying an employee of the farm had to force the lying animals to stand up for the second and third test sessions of ADF.

16.3.4 Data Analysis of the Reliability Study

Measures were calculated for the 6 tests as described in 'Description of test procedures and measures calculated'. The reliability was analysed by calculating Pearson or Spearman rank correlations, depending on distribution of data. Inter-observer reliability was evaluated by comparing paired individual observations, recorded by the experimenter and observer independently at the same time. The results of the two sessions with the two different experimenters were correlated separately, thus there are two values for inter-observer reliability for each test.

To gain information about what sample size is necessary, farm averages from all tested animals from the first test session were correlated with averages where only part of the animals (first 30% of the herd tested or 50% of the herd tested) were taken into account. Also, split-half correlations (correlating the average from the first half of tested animals with the second half) were performed. This was only done with the most promising tests with respect to reliability and validity, i.e., AD and ADF.

16.4 RESULTS

16.4.1 Loose Housing Systems

Validity

Table 16.2 shows the correlations of the test measures with stockpeople behaviour in the milking parlour, based on data of the study by Schmied and Waiblinger (2006a), see also Waiblinger et al. (2007a).

Avoidance distance in the barn (AD) showed the strongest and the highest number of correlations with milker behaviour. The more positive interactions of the milker were observed during routine milking (talking calmly, stroking, and softly touching cows in the milking parlour), the lower was the average AD, the median value of AD on the respective farms and the lower was the proportion of animals with larger AD (> 1m, >2m, >3m). The median value of AD also tended to be lower with a higher percentage of positive handling behaviours.

Table 16.2 Spearman rank correlations between milker behaviour and measures of the animal-human relationship (N=33 for all except WT: N=29).

	Milker behaviour per cow			
	Pos	Neg	Pos%	Neg%
Avoidance distance in the barn (AD)				
AD – farm average	**–.42***	–.05	–.16	.00
AD – farm median	**–.43***	.03	***–.29***	.05
AD – % to touch	.27	.01	.12	–.01
AD – % > 1.0m	**–.49****	.04	–.26	.12
AD – % > 2m	***–.49*****	.02	***–.30***	.03
AD – % > 3m	**–.35***	.01	–.21	.07
Avoidance distance at the feeding place (ADF)				
ADF – farm average	**–.38***	–.04	–.19	–.01
ADF – farm median	–.22	–.14	–.00	–.07
ADF – % to touch	.23	.14	.00	.09
ADF – % > .5m	**–.35***	.13	–.21	.12
ADF – % > 1m	***–.34***	–.18	–.16	–.14
% up Lypass (LP)	–.09	.00	–.13	–.09
% up Lytouch (LT)	–.17	–.28	–.06	***–.33***
% up LP&LT	–.12	–.22	–.01	***–.31***
Walk through and touch (WT)				
WT – % touch	.30	.15	.04	.10
WT – % > 3m	***–.31***	–.11	–.11	–.10
Approach (AP)				
AP –% app contact (of cows standing within 5 m)	–.09	.15	–.11	.21

Notes: Parameters of milker behaviour: frequency: pos = positive, neg = negativ; relative amount: pos%, neg%, = percentage of positive, negative behaviours; correlations with p-values < 0.1 in bold. $^{*}p<0.05$, $^{**}p<0.01$; numbers in italics p<0.1.

Concerning the avoidance distance at the feeding place (ADF), the farm average as well as the proportion of animals with ADF>0.5m correlated significantly and negatively with the number of positive milker behaviour. Moreover, the proportion of animals with ADF>1m tended to correlate negatively with positive handling behaviours.

No significant correlations were found for the percentage of animals standing up during the LP Test. The percentage of animals standing up during the LT test tended to be lower when the proportion of negative interactions was higher.

In the Walk through and touch test (WT), the proportion of animals that withdraw at a distance of > 3m tended to decrease with higher number of positive behaviours.

The approach test showed no significant correlations to milker behaviour.

In the course of the reliability study also inter-test relationships were regarded. High correlations (mostly > 0.7 < 0.9) were found between measures of AD, ADF, and WT (see Windschnurer et al., 2008).

Reliability

In the course of the reliability study a large variation was found in the animal-human relationship on the farms. For instance, the average AD ranged from 0.28 to 1.27m, the median from 0.05 to 1.05 m. Accordingly, the percentage of animals that could be touched ranged from 11% to 68% and the percentage of animals with AD>1.5m varied from 0% to 32% (all for the first test session, regardless the test person). For further descriptive statistics see Windschnurer et al. (2008).

Inter-observer reliability

The inter-observer reliability was very high for all of the tests (Table 16.3). The results with the different experimenter / observer pairs are depicted in scatter-plots in Windschnurer et al. (2008).

Between-experimenter repeatability

As shown in Table 16.6, the approach test (APP), Lypass (LP), Lytouch (LT), and Walking through and touch (WT) showed lower (no to moderate) repeatability as the two avoidance distance tests. At farm level, both avoidance distance measures showed mainly high (≥0.8; bold in Table 16.6) to moderate (≥0.7<0.8; bold and italics in Table 16.6) repeatability. At an individual level, repeatability was still ≥0.7 for both AD and ADF.

TABLE 16.3 Inter-observer reliability of the different tests (at an individual level) Spearman rank correlation coefficients, or % agreement.

	inter-observer reliability		
	Corr.coeff.	p	N
Avoidance distance in the barn (AD) – distance of individual animals	.99/.99[1]	<.001	411/422
Avoidance distance at the feeding place (ADF) – distance of individual animals	.97/.98[1]	<.001	482/497
Lypass (LP) cow up yes/no	agree 100%[2]		501
Lytouch (LT) cow up yes/no	agree 99.8%[2]		455
Walk trough and touch (WT) – individual scores	.96/.96[1]	<.001	130/124
Approach (AP) – individual latencies until contact	1.0/1.0[1]	<.001	87/77

Notes: [1] two coefficients because two people changed the role of being test person plus observer or only observer – both situations calculated separately; [2] see Tables 16.4 and 16.5 for exact description.

TABLE 16.4 Crosstables of observer agreement on cows' reactions during Lypass (LP).

		Lypass person 2		Total
		0	1	
Lypass person 1	0	488.00	0	488
	1	0	13	13
	Total	488	13	501

Notes: 0 = cow stays lying; 1 = cow stands up.

TABLE 16.5 Crosstables of observer agreement on cows' reactions during Lytouch (LT).

		Lytouch person 2		Total
		0	1	
Lytouch person 1	0	359.00	1	360
	1	0	95.00	95
	Total	359	96	455

Notes: 0 = cow stays lying; 1 = cow stands up.

Sample Size

Calculations regarding sample size were only performed for the tests promising due to a high repeatability and validity. During the first test session 72 to 96 percent of the herd were tested (these values are corresponding to 100% tested animals). Correlations of the farm averages calculated from this maximum sample size with smaller sample sizes (30% of the herd, 50% of the herd, first or second half of tested animals corresponding to 36 to 48% of the herd) were high for both AD and ADF (Tables 16.7 and 16.8). Split-half correlations (results of first and second half of tested animals correlated) were lower, but still sufficiently high (Tables 16.7 and 16.8).

TABLE 16.6 Between-experimenter repeatability and test-retest repeatability of the different tests.

	Between-experimenter repeatability: herd & individual level			Test-retest repeatability: herd & individual level		
	Corr. coeff.	p	N	Corr. coeff.	p	N
Avoid. dist. barn (AD)						
AD – distance of individual animals	B .70	<.001	348	B .70	<.001	348
AD – farm average	B .69	.003	16	B .84	<.001	16
AD – farm median	B .81	<.001	16	B .82	<.001	16
AD – % to touch	B .69	.003	16	B .69	.005	16
AD – % > .5m	A .79	<.001	16	A .83	<.001	16
AD – % > 1.0m	B .70	.003	16	B .70	.003	16
AD – % > 1.5m	A .93	<.001	16	A .93	<.001	16
AD – % > 2m	B .48	.059	16	B .52	.041	16
AD – % > 3m	B .51	.045	16	B .70	.003	16
Avoid. dist. feeding place (ADF)						
ADF – distance of individual animals	B .69	<.001	479	B .70	<.001	479
ADF – farm average	A .95	<.001	16	B .92	<.001	16
ADF – farm median	B .85	<.001	16	B .85	<.001	16
ADF – % to touch	A .87	<.001	16	A .87	<.001	16
ADF – % > .3m	A .86	<.001	16	A .87	<.001	16
ADF – % > .5m	B .92	<.001	16	B .88	<.001	16
ADF – % > 1.0m	B .56	.024	16	B .56	.024	16
% up Lypass (LP)	B .31	.246	16	B .48	.061	16
% up Lytouch (LT)	A .59	.017	16	A .59	.017	16
Walk trough and touch (WT)						
WT – individual score	B .64	<.001	47	B .60	<.001	47
WT – median farm score	B .48	.059	16	B .46	.072	16
WT – % touch	B .62	.010	16	B .71	.002	16
WT – % > 1m	A .66	.005	16	A .64	.008	16
WT – % > 2m	B .29	.272	16	A .63	.009	16
WT – % > 3m	B .73	.001	16	B .73	.001	16
Approach (AP)						
AP – average latency to contact	A . 09	.735	16	A .07	.800	16
AP – % app to 2.5m *	A .65	.009	15	A .66	.008	15
AP –% app contact *	A .62	.013	15	A .67	.006	15

Notes: [A] Pearson correlation coefficients; [B] Spearman rank correlation coefficients; * from standing in stable at test start.

TABLE 16.7 Correlations of farm averages of AD (N=16) calculated from 100% of tested animals with the ones calculated from the first and second half of the tested animals, and the first 50% or 30% of the herd that were tested, as well as split-half correlations.

	Average AD 1st tested half	Average AD 2nd tested half	Average AD 50% of herd	Average AD 30% of herd
Average AD 100% tested	r=.94 p<.001	r=.97 p<.001	r=.97 p<.001	r=.86 p<.001
Average AD 1st tested half		r=.82 p=.001		

TABLE 16.8 Correlations of farm averages of ADF (N=16) calculated from 100% of tested animals with the ones calculated from the first and second half of the tested animals, and the first 50% or 30% of the herd that were tested, as well as split-half correlations.

	Average ADF 1st tested half	Average ADF 2nd tested half	Average ADF 50% of herd	Average ADF 30% of herd
Average ADF 100% tested	r=.96 p<.001	r=.88 p<.001	r=.97 p<.001	r=.95 p<.001
Average ADF 1st tested half		r=.75 p=.001		

16.4.2 Tie-stall Systems

Three experimenters performed the tests. Both, between-experimenter repeatability and inter-observer reliability were evaluated at an individual level because only one farm with tied cows was visited. Inter-observer reliability was very high (for the different experimenter combinations: r =0.98/r =0.96/r =0.97, p<0.001, N=54). Between-experimenter repeatability was moderate (for the different experimenter combinations: r=0.54/r =0.46/ r=0.52, p<0.001, N=54).

16.5 DISCUSSION

16.5.1 Validity

With regard to the associations to milker behaviour, quite consistent correlations were found for the two avoidance distance tests (AD and ADF), but no significant relationships were found between milker behaviour and the other tests. The median and average of avoidance distance in the barn (AD) as well as the % of animals avoiding the test person at a distance > 1m or >2m was lower on farms where milkers showed more positive interactions with cows. This confirms earlier results, where a smaller AD was found on farms with a higher frequency and a higher percentage of positive stockperson behaviour (Waiblinger et al., 2002, 2003). However, contrary to findings by Waiblinger et al. (2002), no correlations were found with the overall negative milker behaviour (i.e., sum of talking impatiently, shouting, and forceful hitting with the hand or a stick). Yet, with single variables (e.g. shouting) Schmied and Waiblinger (2006a) found significant correlations. In this study farms were selected according to handling practices during rearing, while in the earlier studies farms were selected randomly. Handling practices during rearing were linked to later responses of cows to humans (Schmied and Waiblinger, 2006b; Waiblinger et al., 2007a). They are likely to interact with daily human–animal interactions with regard to effects on the animals' reactions to humans. This may explain the lack of correlation with negative milker behaviour. Furthermore, due to time constraints, milkers filled in an attitude questionnaire about behaviour towards cows directly before being observed in the milking parlour. Although they were told that the cows' behaviour was observed, milkers may have been more reluctant to use negative behaviour after being sensitised by the questionnaire.

The average of the avoidance distance at the feeding place (ADF) and the percentage of cows withdrawing at a distance of at least 0.6 m in ADF were also lower on farms where milkers showed more positive interactions. But correlation coefficients were lower compared to measures of AD. This is also in line with a previous study by Waiblinger et al. (2003), who found that ADF was more influenced by possible confounding factors.

No correlation was found between milker behaviour and the animals' reactions in the approach test (APP). Although we would have expected a lower correlation with milker behaviour compared to avoidance distance due to previous results (Waiblinger et al. 2003, Waiblinger and Rousing 2004), we still had expected some kind of significant correlation. Possible factors contributing to this result might be (i) the pre-selection of farms, (ii) the modified test situation (taking into account not all standing animals as done in Waiblinger et al. 2003, but restricting it to a special distance) or (iii) differences in test time.

For the new and partly simpler tests, the Walk through and touch test (WT) and the Lypass and Lytouch (LP and LT) tests, no significant correlations were found with milker behaviour, although a tendency was found for WT.

In the reliability study, where convergent validity of the behavioural tests was investigated by means of inter-test relationships, high correlations were found between measures of AD, ADF, and WT (Windschnurer et al., 2008).

In sum, the relatively high correlations with milker behaviour again confirm the validity and relative robustness of AD to measure the cows' relationship to humans and thus say something about the level of negative or positive emotions during interactions with humans as well as chronic stress. ADF also has some validity, but – in agreement with earlier results (Waiblinger et al., 2003) – seems to be more influenced by other factors. Thus, it might reflect the cow-human relationship to a lesser extend. With regard to the new test WT, it seems promising because of its high convergent validity when compared with AD and ADF (Windschnurer et al., 2008), although the measures of the test only tended to correlate with milker behaviour. At the moment, the other tests seem not to be valid, although further investigations may be necessary to confirm this finding under different farm conditions.

16.5.2 Reliability

Loose Housing Systems

The results show that for all tests a very high inter-observer reliability can be achieved. For the simple tests Lypass (LP) and Lytouch (LT), where the reaction is just noted dichotomously (standing up: yes/no), this was expected. When performing the avoidance distance tests (AD and ADF) as well as the Walk through and touch test (WT) not only the reaction of the animal has to be judged correctly (withdrawal or not), but also the distance has to be assessed. The results show that a very high reliability can be reached with sufficient training, where devices to ease distance assessment are learnt.

Concerning between-experimenter repeatability (which is also influenced to some extend by a possible error of assessment by two different persons, i.e., inter-observer reliability) at farm level only AD and ADF gained sufficiently high values, reaching correlation coefficients >0.8 and up to >0.9 for several measures. Consistently high repeatability was found for instance for the median AD and ADF. The between-experimenter repeatability at an individual level, i.e., the individual animals' consistency in reaction to different unknown people, was still moderately high ($r \geq 0.69$) for both AD and ADF. This is remarkable high, as test repetition or small differences in the test persons' behaviour are known to be confounding factors (Waiblinger et al., 2006). The experimenters in our study were apparently able to standardise their behaviour quite well. For Lypass (LP) and Lytouch (LT) the somehow restricted sample (lying animals), but also the variability in the farms studied may have caused lower repeatability. A better repeatability could be gained when the tests were combined to one test (Windschnurer et al. 2008), probably due to the low variation in responses of animals during LP. For WT only moderate ($r < 0.7$) repeatability was reached. Both WT and the approach-test (APP) test only a restricted number of cows, being in the respective area at the special time. This may cause the lower repeatability. Altogether, compared to AD and ADF, the other tests are more difficult to standardise with respect to sample size and sample selection, which may cause the lower repeatability. Nevertheless, in our opinion the WT test merits further investigation in future, as it may be interesting especially in larger herds. However, this has to be investigated. For more detailed discussion regarding inter-observer reliability and between-experimenter repeatability including potential observer bias see Windschnurer et al. (2008).

We also calculated the repeatability of the cows' reaction when tested twice. However, due to our study designed to investigate between-experimenter repeatability, the potential inconsistency in reaction to different people contributes to possible differences and vice versa. Test-retest repeatability was mostly very similar or even equal to the between-experimenter repeatability, thus, not allowing a differentiation between effects of person or test repetition.

The results also suggest that smaller sample sizes may be sufficient to reflect the animals' relationship to humans of the whole herd.

Tie-stall Systems

Only one tie stall farm could be visited in the course of the reliability study. Three experimenters subjected 54 animals each to the avoidance distance test at the feeding place (ADF). Inter-observer reliability was comparable to ADF in loose-housed cows, but the between-experimenter repeatability was lower. However, also in loose-housed animals the between-experimenter repeatability was lower at an individual level. The lower between-experimenter repeatability compared to ADF in loose housing systems might have been partly caused by the fact that after the first test session (=cows tested by the respective first experimenter), quite a lot of animals were lying and were forced to stand up by a human, enlarging partly the distance in the second and third test session.

In a previous study including 9 tie stall farms, Waiblinger et al. (2007b) evaluated the between-experimenter repeatability of an avoidance test at the feeding place using an 11-point score, with 2 to 3 experimenters out of 4 performing the tests in a balanced order.

Low to high (average correlations ranging from 0.37 to 0.88) between-experimenter repeatability was found within farms, i.e., at an individual level. At farm level, between-experimenter repeatability was moderate to high (0.65 to 0.80) and thus promising. The substantial differences of between-experimenter repeatability within farms were attributed to partly inconsistent reactions towards different experimenters or subtle differences in the experimenters' behaviour. Taking these previous results into consideration we would expect to be able to reach sufficient between-experimenter repeatability at farm level when performing the test like in the present study, using a metric scale.

16.6 CONCLUSIONS

Measuring avoidance distance in the barn (AD) proved to be not only valid, but also highly reliable. The latter also accounts for the avoidance distance at the feeding place (ADF), where less strong evidence was found for its validity when correlated with milker behaviour. Nevertheless, our results suggest that both can be recommended for inclusion in a welfare assessment scheme. Each measure has its advantages and disadvantages. For validity reasons we would prefer AD since it seems to be less influenced by confounding factors compared to ADF. With regard to the new test Walk through and touch (WT), it seems promising because of its high convergent validity when compared with AD and ADF, although the measures of the test only tended to correlate with milker behaviour.

17

ASSESSMENT OF HUMAN–ANIMAL RELATIONSHIPS IN FATTENING BULLS

I. Windschnurer, X. Boivin and S. Waiblinger

17.1 SUMMARY

The aim of this project was to test different tests for measuring the Animal Human Relationship in fattening bulls with respect to their feasibility and reliability.

First, new tests were developed for fattening bulls and pre-tested on 6 farms, ending up with the avoidance distance at the feeding place (ADF) – comparable to the test performed with dairy cows - being the only feasible and seemingly valid test. Repeatability (between-experimenter and test-retest) was tested on 10 farms, while inter-observer reliability was tested on 6 out of the 10 farms. Inter-observer reliability, analysed at an individual level, was very high (r_s = 0.96 / 0.97, p< 0.001, N=288 / 297), and between-experimenter repeatability sufficiently high, especially for the larger distances, i.e., the more fearful animals (average farm ADF: r_p =0.76; ADF % >0.5m: r_p =0.77, p<0.05, N=10).

17.2 INTRODUCTION

Recent preliminary results suggest that also in fattening bulls the animal's relationship to humans (AHR), based on their previous experiences with humans, has an important influence on animal welfare and productivity (Probst et al., 2008). To our knowledge, there exist no simple and quick tests for assessing the animal's relationship to humans (AHR) in fattening bulls on-farm. In the literature, evaluations of temperament and docility of bulls can be found, but they often include handling / manipulation of the animals (for a review see Burrow, 1997). Several handling tests were described for male cattle measuring their responses when restrained in handling facilities and confronted with close human contact (for an overview see also Windschnurer et al., 2009). Tilbrook et al. (1989) investigated responses of young bulls and steers to humans in an arena test, where the approach behaviour to a stationary test person was recorded individually. Comparing measures of temperament, Fisher at al. (2000) worked with a paddock and yard flight distance test where a human approached stationary steers until they moved. However, for an over-all welfare assessment protocol with time limits it would not be feasible to construct an arena or spent a lot of time separating animals in order to test them individually. Our aim was to find feasible and simple tests for assessing the AHR in fattening bulls and to evaluate these tests with regard to their reliability (inter-observer reliability, between-experimenter and test-retest repeatability, see Chapter 16 for definitions). Therefore, tests for assessing animal-human relationship that have been used in dairy cows, such as approaching animals in order to test avoidance reactions or an approach test to a stationary unfamiliar human (Waiblinger et al. 2003, Rousing and Waiblinger, 2004) were evaluated according to their feasibility during visits to test farms. Attempts were made to modify the tests for the use on bull fattening farms.

17.2.1 Development and Choice of Tests

In order to pre-test different tests and measures 6 farms were visited. Several potential tests were checked for their feasibility and potential to validly reflect the bulls' relationship to humans, including fear of humans, in different housing systems.

When developing or modifying tests for fattening bulls, one constraint is that the tests must be performed from outside the pen because entering pens would be too dangerous. Moreover, interacting with the bulls too closely can be dangerous, even from outside the pen. Thus, (1) an approach test with the stationary test person located close to the (post-and-rail) feed barrier and (2) an avoidance distance test, both being tests that have been performed in some way with dairy cows before, were evaluated for their feasibility and their potential validity.

(1) Several modifications of an approach test were tested, varying a number of factors, such as the test duration, the exact procedure (e.g., standing the whole time in one place or changing places in a specified time schedule to allow more animals to approach to contact) and the behaviour of the test person (e.g., adding moving elements such as trying to touch and stroke bulls that had approached close enough). However, none of these modifications resulted in a feasible and seemingly valid test because of the following problems:

- not all animals seemingly interested in making contact to the experimenter could gain access because of (1) dominant animals or (2) animals lying directly at the feeding place hindering them
- the more complex variants were too difficult to perform in a reliable way

(2) We only slightly modified the avoidance distance test at the feeding place which has been previously used in dairy cows, abandoning longer duration of stroking to avoid dangerous situations when bulls are grasping the experimenter with their mouth. The

avoidance distance test at the feeding place proved to be a simple and quick test that can be easily applied on-farm under different housing conditions, also with fattening bulls. However, feeding times have to be taken into account in order to have a sufficient number of animals at the feeding place.

In sum, we decided to further evaluate the avoidance distance test, which is described in detail in Windschnurer et al. (2009).

17.3 METHODS

17.3.1 Farms, Animals, and Housing

Conducting an on-farm survey, 10 farms with altogether 123 pens (housing 3 to 12 bulls, on average 6 animals per pen) were tested. Farms had between 8 and 18 pens and 40 to 123 bulls (73±24). On one farm bulls were housed in deep litter, on 9 farms bulls were housed on fully slatted concrete floors (except for two pens with straw bedding). The weight classes ranged from estimated 200kg up to finishing bulls with approximately 700kg. The animals were mostly Simmental bulls, the breed traditionally found on commercial Austrian bull fattening farms.

17.3.2 Design of Reliability Tests

The avoidance distance test (ADF) was always performed by the same two trained test persons during a one-day visit to the farm. The test persons (female, both 1.68m) always wore green overalls. The average weight of the bulls per pen was assessed before performing the test to be able to balance the test order of pens for weight classes. The age of individual bulls was noted from farm records after the tests were finished.

Repeatability was assessed on all farms, while inter-observer reliability was evaluated on 6 of the 10 farms. On the farms where inter-observer reliability was studied, the two test persons changed their role (being experimenter or just observer) with each pen. After having tested all pens in such a way (= first test session), they started re-testing the pens so that the animals were tested by the other person who had been observer for the respective pen before. On the farms where inter-observer reliability was not studied, the pens within each farm were assigned to two groups, taking the weight classes into account. One half was tested by one experimenter, while the other pens were tested by the second experimenter at the same time. After a crossover, each experimenter tested the pens tested by the other person during the first session. Both experimenters tested the same amount of pens of the same weight classes during both test sessions.

After performing the ADF test twice with all of the pens a novel object test was carried out additionally with part of the pens. This additional test was designed to assess general fearfulness (see Chapter 22).

17.3.3 Procedure

The test is comparable to the ADF test performed with dairy cows (Waiblinger et al., 2003), described and pictured in chapter 16, where a moving human approaches individual animals. It was performed during the main feeding times, starting 5 minutes after food was distributed. After choosing an animal, it is approached from the front, one arm held in an angle of 45° in front of the body, with a speed of 1 step per second starting from a distance of 2m, if possible. The animal is approached until it withdraws or until touching. If the bull allows the touch on the muzzle, the test person tries to stroke its cheek for at least 1s but not over 3s. When the animal reacts by showing clear signs of avoidance or withdrawal, the distance between the test person's hand and the muzzle / nose is estimated at the moment of withdrawal in steps of 10 cm. If an animal can be approached until touching or stroking the avoidance distance is 0m.

17.3.4 Data Analysis

Pen average, farm average and farm median as well as at farm level the percentage of animals with certain distances (ADF = 0m; >0.2m; >0.3m; >0.5m) were calculated. Since the experimenters tested not always the same animals and we wanted to investigate the possible influence of individual animals, the measures farm average and farm median were calculated for all animals tested as well as only for animals tested by both experimenters.

Inter-observer reliability and repeatability (between-experimenter and test-retest) were analysed by means of Pearson and Spearman rank correlations, depending on the distribution of the data. Inter-observer reliability was analysed at an individual level since individual observations were compared. With regard to repeatability, individual, pen and farm measures were evaluated. For the calculations at pen level, only animals that were tested by both experimenters were included due to the low number of animals per pen, which reduced the sample to 122 pens. That means, pen values were sensitive to values of individual animals. At farm level, repeatability was calculated including all animals tested as well as only the ones tested by both experimenters.

In order to study a possible effect of age or weight (1) correlations were calculated between avoidance distances and age of individuals (based on data from two farms). Moreover, (2) avoidance distances from the three weight classes 1 (< 300 kg; N=138 / 95 in the first / second test session), 2 (>300–500kg, N=265 / 190) and 3 (>500kg–700kg, N=256 / 183) were compared using the Kruskal-Wallis test (performed with data obtained from all 10 farms). If bulls were tested the first time during the second test session, their measure was put to the measures of first test session (ADF1) with regard to this weight effect analysis.

17.4 RESULTS

17.4.1 Inter-observer Reliability

Inter-observer reliability was very high regardless which person was the experimenter or observer (r_s = 0.96 / 0.97, p< 0.001, N= 288/297). When excluding animals which could be touched or stroked, inter-observer reliability was still high (r_s = 0.96 / 0.93, p< 0.001, N= 169/125).

17.4.2 Repeatability

Table 17.1 shows the repeatability for the different measures.

At farm level partly high (≥0.8) or moderate (≥0.7<0.8) correlations were found for between-experimenter repeatability and test-retest repeatability of the measure average avoidance distance (taking all animals or only animals tested by both experimenters into account). No sufficient repeatability was found for the measure median avoidance distance, whereas for the other farm measures, repeatability was moderate to high (0.4–0.7 and > 0.7, respectively) for animals tested by both experimenters (e.g., for ADF % >0.5 m, r_p= 0.79).

At pen level, only low to moderate (< 0.4 and 0.4–0.7, respectively) correlations were found (e.g., for the measure pen average: between-experimenter: r_p=0.61, p<0.001, N=122, test-retest: r_p=0.64, p<0.001, N=122).

Also at an individual level only moderate correlations were found (between-experimenter: r_s=0.58, p<0.001, N=469, test-retest: r_s=0.58, p<0.001, N=469).

17.4.3 Consistency in Bulls' Reaction: Effects of Experimenter and Repeated Testing

The avoidance distance of individual bulls decreased from the first to the second test (Z= –4.37, p = 0.000, N=469; Figure 17.1 left). Accordingly, in the first test (ADF1) the farm average was 0.15cm and the median ADF1 was 0.10m, whereas the average ADF2 was 0.12m and the median ADF2 was 0.05m.

Moreover, a slight experimenter effect was found: The bulls showed significantly shorter avoidance distance towards the test person B (Z= –3.025, p=0.002, N=469) compared with person A (Figure 17.1 right). This result seems to rely mainly on the very short distances of ≤ 0.2 m (Table 17.2).

TABLE 17.1 Between-experimenter repeatability and test-retest repeatability of ADF at an individual, pen or farm level.

	Between–experimenter repeatability: farm, pen & individual level			Test–retest repeatability: farm, pen & individual level		
	corr.coeff.	P	N	corr.coeff.	P	N
Avoidance distance at the feeding place (ADF)						
ADF – distance of individual animals	[B] .58	<.001	469	[B] .58	<.001	469
ADF (bulls with ADF> 0m)	[B] .43	<.001	154	[B] .43	<.001	154
ADF – pen average	[A] .61	<.001	122	[A] .64	<.001	122
ADF – pen median	[A] .49	<.001	122	[A] .47	<.001	122
ADF – farm average*	[A] **.87**	.001	10	[A] **.81**	.005	10
ADF – farm average**	[A] **.76**	.011	10	[A] .63	.051	10
ADF – farm median*	[B] .25	.482	10	[A] .67	.033	10
ADF – farm median**	[A] .32	.367	10	[A] **.71**	.021	10
ADF – % to touch*	[A] .54	.108	10	[A] .61	.062	10
ADF – % > .2m*	[A] **.71**	.021	10	[A] .66	.038	10
ADF – % > .2m**	[A] .67	.035	10	[A] .63	.053	10
ADF – % > .5m*	[A] **.79**	.006	10	[A] .65	.043	10
ADF – % > .5m**	[A] **.77**	.009	10	[A] .41	.235	10

Notes: [A] Pearson correlation coefficients; [B] Spearman rank correlation coefficients; coefficients ≥ 0.7 in bold; * only animals tested by both experimenters, pairs; ** all tested animals, even if only tested once by one of the two experimenters.

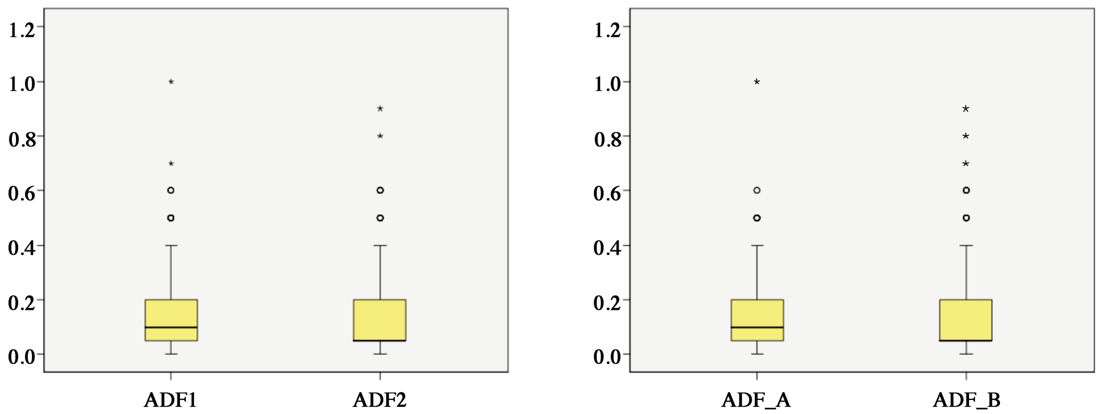

FIGURE 17.1 Box-whisker plots of the individual ADF of the same bulls (N=469) in the first (ADF1) and second test (ADF2) (left) or tested by person A (ADF_A) or B (ADF_B) (right).
Notes: Right graph modified from Windschnurer et al., 2009.

17.4.4 Effects of Confounding Factors (Age or Weight)

ADF showed significant correlations with age (Table 17.3). Accordingly, the three weight classes differed significantly in the first (chi-square=15.02, p=0.001) and second (chi-square=27.02, p<0.001) test session (Figure 17.2). With regard to the first test session, significant differences were found between group 1 and 2 (Z= –2.80, p=0.05) and group 1 and 3 (Z= –3.85, p<0.001), whereas there was no significant difference between group

TABLE 17.2 Overview of the distribution of the categories of ADF on the 10 bull fattening farms for the two test sessions and the two test persons.

	stroke	AB	Touch	> 10cm	> 20cm	> 30cm	> 50 cm
Test person A							
N	10	10	10	10	10	10	10
Average	6.26	40.48	46.73	35.27	16.88	7.43	0.72
Minimum	0.0	26.2	29.5	13.2	2.6	0.0	0.0
Maximum	16.3	54.1	58.5	53.2	30.2	13.2	2.7
Test person B							
N	10	10	10	10	10	10	10
Average	10.66	48.69	59.34	26.33	15.25	9.00	2.31
Minimum	4.9	32.4	42.1	7.9	.0	.0	.0
Maximum	16.2	64.7	76.5	44.7	29.8	21.3	8.1
1st Test							
N	10	10	10	10	10	10	10
Average	6.02	41.31	46.2	36.00	21.5	8.64	1.28
Minimum	0.0	32.8	32.9	13.2	3.6	.0	.0
Maximum	13.2	54.7	59.5	55.3	35.1	17.0	3.8
2nd Test							
N	10	10	10	10	10	10	10
Average	10.89	47.84	58.7	25.35	13.4	8.05	1.84
Minimum	6.4	37.2	46.8	7.9	.0	.0	.0
Maximum	14.0	70.6	82.4	38.3	23.3	14.9	8.1

Notes: stroke = animal allows stroking of the cheek, AB = avoiding at the moment of being touched; touch = stroke plus AB = avoidance distance 0m.

TABLE 17.3 Spearman rank correlation coefficients for the avoidance distances of individual bulls on two farms.

Spearman – Rho	ADF1	ADF2	ADF_ exp. A	ADF_ exp. B
Age	$r_s = -0.26$	$r_s = -0.29$	$r_s = -0.36$	$r_s = -0.21$
	$p = 0.004$	$p = 0.007$	$p < 0.001$	$p = 0.033$
	N= 119	N= 84	N= 103	N= 100

Notes: ADF1/2: first or second test; ADF_exp.A/B: ADF to experimenter A or B.

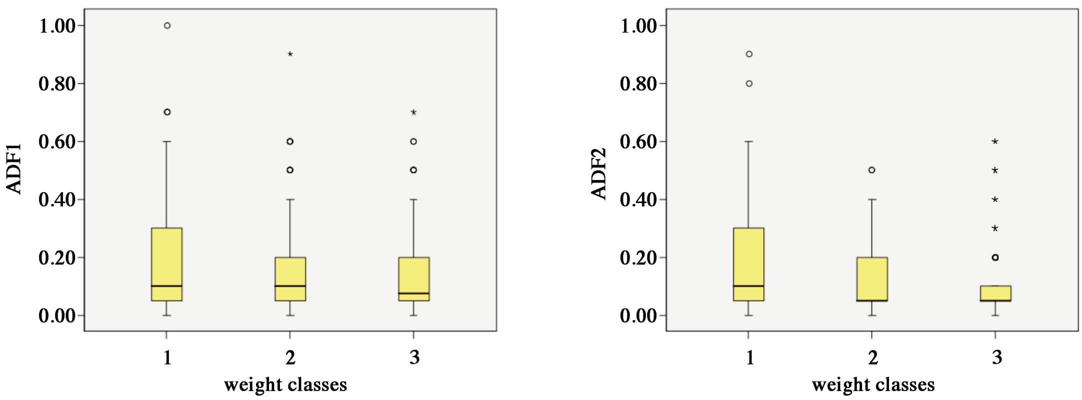

FIGURE 17.2 Box-whisker plots for ADF1 (first test session) and ADF2 (second test session) of the three weight classes 1 (<300kg; $N_{ADF1/ADF2}$=138 / 95), 2 (>300–500kg, N=265/190) and 3 (>500kg–700kg, N=256 / 183).

2 and 3 (Z= –1.26, p=0.206). With regard to the second test session, significant differences were found between all three weight classes (group 1 and 2 (Z= –2.51, p=0.012), group 1 and 3 (Z= –4.87, p<0.001), as well as group 2 and 3 (Z= –3.40, p=0.001).

17.4.5 Sample Size

When correlating the average farm ADF, based on 100% of the pens on the farm, it highly correlated with measures based on a reduced sample size, taking only animals from the first 50% of pens or the second 50% of pens into account (r = 0.94/ for 1st 50%; r=0.90 for 2nd 50%, p<0.001, N=10). Only values from animals tested by both experimenters were included into this first calculation.

17.5 DISCUSSION

The avoidance distance test for assessing the reactions of fattening bulls towards humans turned out to be easy and quick to perform when taking feeding times into account.

Similarly to findings in dairy cows (Windschnurer et al., 2008), the inter-observer reliability was very high, also when excluding animals that could be touched, i.e., where no distance has to be assessed and thus low to no disagreement in estimation of distances can be expected.

Repeatability at an individual level and at pen level was moderate (mostly r~0.6) and lower than found between-experimenter repeatability at an individual level in dairy cows (r_s=0.69) (Windschnurer et al., 2008; Chapter 16). At farm level, the between-experimenter repeatability was still somewhat lower and less consistent compared to what we found in dairy cows, but the farm average still showed high repeatability as did the percentage of animals with ADF >0.5m, i.e., the more fearful animals. The repeatability in the more fearful animals seems to be higher, which is even more important with regard to welfare assessment since extremes should be detected. Thus, measures reflecting higher distances should be used for the welfare score, e.g., by using the farm average, which better reflects extremes than the median value, or the percentage of animals with ADF>0.5m or 0.2m. The latter may be more appropriate to detect variation in farms with similarly small avoidance distances, such as the farms in our study.

The higher repeatability of higher distances is also reflected in the experimenter bias we found, i.e. bulls showed less avoidance towards one of the two experimenters. Experimenter B could touch more animals and approach more animals up to a distance of 10 cm compared to experimenter A. Although this may have contributed to the somewhat lower repeatability compared to dairy cows, it has to be kept in mind that the differences were small and mainly due to the animals with very small ADF of below 0.2m, thus supposedly biologically meaningless. Also in dairy cows, a directed experimenter effect was found (Waiblinger and Menke, 2003). However, it was confounded with the order of testing, since always the same person performed the first test. Subtle differences in the behaviour of the two experimenters during testing may have contributed to the difference found in the present study (Waiblinger et al., 2006).

Reactions to an approaching human in general were found to be valid in measuring the animal's relationship to humans in several species including cattle (Waiblinger et al., 2006). Accordingly, the avoidance distance at the feeding place was suggested to be a valid measure of AHR in dairy cattle (Waiblinger et al., 2003; Windschnurer et al., 2008, Chapter 16). A preliminary evaluation of the validity of this test for fattening bulls is described in Windschnurer et al. (2009). Avoidance distances of fattening bulls were shown to be partly linked with farmers' attitudes.

We found a significant weight or age effect with the lighter or younger animals being more fearful and thus having higher avoidance distances. The variability between farms was lower for the heaviest animals. Therefore, the distribution of weight classes has to be taken into account when comparing different farms and, if selection is necessary, preferably the very heavy animals should be disregarded. As well, the inclusion of animals that recently arrived on the test farm has to be avoided, because it is likely that the results might rather reflect previous experiences on other farms than the actual human-animal interactions on the test farm. Thus, it is important that at the time of performing the tests, the animals should have spent some time on the farm. Furthermore, inquires into this direction (.e.g., the necessary duration of being on the farm) would be necessary. However, test results of these young (recently arrived) animals still can be used to make people aware of potential problems regarding the AHR, independently from the source of the problem. The farmer could be advised to act, like providing additional positive contact (Windschnurer et al., 2009).

The reason for decreasing avoidance with increasing age / weight are not clear and merit further investigation as this would also add information about how the older bulls perceive the humans and thus about the validity of the test.

17.6 CONCLUSIONS

Taken all results together, the ADF test seems to be a promising test to include into welfare assessment for fattening bulls. It is easy and quick to perform when taking feeding times into account and the repeatability is quite high at farm level. Confounding factors (weight, age) need to be considered during the assessment. For the moment, it seems to be the only possibility – feasible for on-farm assessment – to gain information about the bulls' relationship to humans and thus the level of negative or positive emotions experienced during interactions with humans and probably also chronic stress. Results in dairy cows and first results in fattening bulls suggest that it could be a valid measure of AHR. Further investigations with respect to the validity of the test for fattening bulls are necessary.

18

VALIDATION OF EXPLORATORY BEHAVIOUR IN BEEF CATTLE AS AN INDICATOR OF POSITIVE EMOTIONS

H. Schulze Westerath, N. Brörkens, S. Laister, N. MacKintosh, C. Winckler and U. Knierim

18.1 SUMMARY

In the past, usually the assessment of animal welfare focussed on indications of impaired welfare. During the last years, however, the interest in indicators of good welfare, namely positive emotions, has increased. Explorative behaviour, and especially intrinsic exploration which is directed towards stimuli which may have no biological significance, appears to be performed for its own purpose. In this respect it is very similar to play behaviour. Animals typically show a high motivation to explore which suggests that it is perceived intrinsically pleasant. At the same time, the acquisition of information may increase predictability and controllability, two criteria that are regarded important for maintaining good welfare. It was our goal to develop and validate a feasible on-farm test to assess the level of exploratory behaviour that beef bulls are stimulated and able to perform in their daily farm environment. We assumed that animals kept in barren environments have less stimulation and ability to perform exploration than animals in enriched environments, and when confronted with a novel object would show more or prolonged interest, as it has been shown in pigs and rats already. After selection of a suitable test object (cross of two green plastic hosepipes, 1.5 cm in diameter, 40 cm long) and setting up of the test protocol, we compared the average time bulls were in contact with a novel object with different behaviours in barren fully slatted systems with bulls in deep litter systems. As many novel objects as feeding places or bulls were provided, and the tests lasted for one hour. Sixty one groups of bulls weighing between 350 and 550 kg on 19 farms in Germany and Austria were investigated. Additionally, on 5 of the experimental farms with barren pens half of the groups (11 groups) were given simple enrichment for one week after which the test was repeated in all experimental groups on these farms.

The objects were immediately approached in all groups. Differences in the exploratory behaviour of bulls kept in different housing systems were found to some extent. When calculated per animal, licking/chewing, horning/rubbing as well as the total occupation

with the objects was significantly less frequent in deep litter systems compared with slatted floor pens. Interpretation of the results was difficult because of partly different animal to feeding place ratios in littered versus slatted pens. However, later analysis of a complemented data set with nearly similar animal to feed place ratios (Schulze Westerath et al., 2009) confirmed significant, but only slight differences for total occupation with the objects as well as licking/chewing. In general, total occupation with an object decreased over time with markedly higher values in the first 15 minutes and no differences between the third and fourth quarter of the test. Furthermore, no general effect of the short-term enrichment on any of the measures could be detected. Results of the repeated tests suggest that some degree of habituation occurred although total time in contact with the object did not decrease. At the current stage of development no feasible and valid test of exploration in beef cattle can be recommended. Questions certainly deserving more investigation in the future are the actual exploratory behaviour in relation to test results and the motivational background of different exploratory behaviours.

18.2 INTRODUCTION

In the past, usually the assessment of animal welfare focussed on indications of impaired welfare. During the last years, however, the interest in indicators of good welfare, namely positive emotions, has increased (e.g. Knierim et al., 2001; Désiré et al., 2002). The search for such indicators is at its early beginnings and has not yet generated many candidate indicators, especially not those that are suitable for use in an on-farm welfare assessment (Winckler et al., 2003).

Explorative behaviour, at least ‘intrinsic exploration’ which is directed towards stimuli which may have no biological significance (Wood-Gush and Vestergaard, 1989) is behaviour that appears to be for its own purpose (Fowler, 1965; c.f. Murphy, 1978). In this respect it is very similar to play behaviour. Animals typically show a high motivation to explore which suggests that it is perceived intrinsically pleasant. At the same time exploration allows to gain information about the environment which may increase predictability and controllability, two criteria that are regarded important for maintaining good welfare (Wiepkema, 1987). When exploration is shown a situation must be assessed as novel, and fear must necessarily be limited (Murphy, 1978). Boredom and apathy have been discussed as possible consequences of deprivation from intrinsic exploration (Wood-Gush and Vestergaard, 1989; Wemelsfelder and Birke, 1997). On the basis of these considerations, in this study we presupposed that exploratory behaviour is associated with positive emotions.

It was our goal to develop and validate a feasible approach for the on-farm assessment of the level of exploratory behaviour that beef bulls are stimulated and able to perform in their daily farm environment. As it is problematic to reliably record levels of daily exploratory behaviour in short-term observations because of their sporadic, irregular and context-dependent nature, we applied a further presumption, namely that animals being stimulated and able to only perform low regular levels of exploratory behaviour will show increased or prolonged responses to a novel object compared to animals that regularly perform higher levels of exploratory behaviour as already described for pigs (Wood-Gush and Vestergaard, 1989) and rats (Zimmermann et al., 2001). Therefore, the use of a novel object test appeared to be an option, where less exploratory behaviour in the test would mean that more exploratory behaviour is performed in the animals' daily life and, hence, the more positive emotions they would experience. Cattle exploratory behaviour concerning a novel stimulus has been investigated in cows (Herskin et al., 2004; Schrader, 2002), heifers (Boissy and Le Neindre, 1990; Boissy and Bouissou, 1995; Hemsworth et al., 1996; Kilgour et al., 2006), bulls (Hemsworth et al., 1996), steers (Kilgour et al., 2006) and calves (Miller et al., 1986; Dellmeier et al., 1990; Van Reenen et al., 2004, 2005). Except for one study with cows in tie-stalls (Herskin et al., 2004) and one in loose housing (Schrader, 2002), all tests were done with novel objects in an arena in social separation. However, to be suitable as an on-farm-test, it must be practicable in the home pen of group housed animals within a reasonable amount of time and expenditure of material. Moreover, it must be workable by one assessor without entering the pen (for safety reasons).

It was decided to develop the test for bulls between 350 kg to 550 kg as this weight class can most regularly be found on bull farms and they also appeared to have the highest potential to be included into the assessment of further welfare measures.

In order to validate the test, we further assumed that animals kept in barren environments have less stimulation and ability to perform exploration than animals in enriched environments. We therefore compared the test results from bulls kept in barren pens with fully slatted floors to bulls kept in deep litter pens. For further validation, we provided bulls in barren pens with simple enrichment devices for one week and compared their test results with animals from unenriched pens at the same farm in order to investigate possible effects of a short and simple enrichment on the exploratory behaviour in the test.

18.3 DEVELOPMENT OF AN EXPLORATION TEST

18.3.1 Methods

First investigations into possible types of novel objects, ways of presenting them, timing of the test and possible behavioural measures were carried out with 14 groups of bulls (4 10 animals per group) on four German farms on different times of the day. Three farms had pens with fully slatted floors (SF), one had SF pens as well as deep litter pens (DL). Three different non-edible objects were tested for responses of the animals as well as for their durability:

- a plastic flower-pot, 12 cm high, 15 cm in diameter at the top, wall about 5 mm thick, fixed at the bottom with a metal chain, 15 cm long, hanging inverted at the feeding fence like a bell;
- a plastic chain (cordon) with a red and a white string (each 22 cm long, chain-links: 5 cm long, 2 cm wide), fixed with a snap hook at the feeding fence;
- a cross of two green plastic hosepipes (1.5 cm in diameter, 40 cm long), fixed with a snap hook at the crossover of the strings for fixation at the feeding fence.

In different sessions, the animals were either confronted with objects of one type or with objects of all three types simultaneously (resulting in a kind of choice test) and observed for 30 min or 1 h.

18.3.2 Results

In general, latency to first contact with any of the objects was about 0 seconds. Behaviours that were shown in contact with the objects were sniffing, licking, chewing, horning, rubbing and jostling (Table 18.1). Least occupation was seen with the plastic chain. At the flower pots some animals showed increased horning that appeared to be aggressive. At the hosepipe-cross the animals showed the greatest diversity of behaviours. The flower pot was destroyed once and the hosepipe-crosses four times, however, the construction of the hosepipe could be sufficiently improved during the pilot studies. On this basis, the hosepipe-cross was assessed to be the most suitable object for the test (Figure 18.1).

TABLE 18.1 Behaviour performed in contact with novel object.

Name	Description
Sniffing	Muzzle directed to object, often slight movements of the nostrils, distance to the object 0 to 5 cm
Licking or chewing	Contact with tongue or inner side of the mouth with object or fixating device
Jostling	Contact with the object and the area between muzzle and eyes, object being moved
Horning or rubbing	Area at the head above the eyes having contact with the object, additionally rubbing movements or repeated bumps against the object

FIGURE 18.1 Objects fixed at a feeding rack or neck rail.

A presentation of the objects at the height of the animals' heads, i.e. at the feeding fence, ensured that all animals were able to see the object and to make contact with it (Plate 1). The fixation of the objects at the feeding fence is feasible and takes about one to two minutes. No serious aggression between bulls was observed when one object per feeding place or animal (in case that there are more feeding places than bulls) was provided. If the feeding place consisted of a simple neck rail, 70 cm were taken as one feeding place. There were no indications that the presence of the observer on the feeding table, when standing quiet, calm and about 2 m away from the feeding gate as well as not directly in front of the experimental groups, did disturb the animals.

The highest response to the test in terms of number of animals involved was during periods of general activity and when most bulls were not feeding. Therefore, a suitable test time is a time-window of two hours, starting one hour after feeding (i.e. either provision of fresh food or pushing remaining food towards the feeding fence with or without feeding concentrate) in the morning.

It was possible to test up to 3 groups simultaneously within one session, depending on location, visibility and size of the groups. A maximum number of about 15 objects, either in one group or in up to 3 groups could concurrently be observed. A recording interval of 30 seconds turned out to be a reasonable interval and was applicable with the chosen behavioural parameters. Within one hour of observation, a decline in occupation with the objects could be noted, which indicates that this observation time is sufficient to determine possible differences in the duration of interest in the novel object.

18.4 VALIDATION OF EXPLORATION TEST

18.4.1 Methods

Sixty one groups of beef bulls were tested in ten farms each in Germany and Austria. Animals were intact males of different breeds and were housed in fully slatted floor pens or in deep litter systems in half of the farms each. Estimated body weights were between 350 kg and 550 kg.

The tests were conducted by one person in Germany and one in Austria. First, strings for hanging the novel objects at the feeding gates or rails were fixed at all pens that were observed on the farm. The test procedure started with the fixation of objects at the feeding fence of the selected groups (up to 3 groups) one hour after feeding in the morning. Immediately afterwards the groups were observed by instantaneous scan sampling every 30 seconds for 1 h in total. The numbers of animals showing the behaviours listed in Table 1 were recorded, and additionally the numbers of those lying, feeding (muzzle in contact with food or chewing with head above food) and being near to an object without occupation (head above food or within a head length from an object, no drinking or other behaviour listed above). On all but one farm, a second test session with other groups started about 15 min after the end of the first, and was conducted in the same way.

After the tests, half of the group pens of the German farms with slatted floors were 'enriched' with two strings of a metal chain (about 50 cm and 25 cm long) and two strings of a rope (each 20 cm long), fixed at the pen boundaries at the height of the animal's heads. The objects were left in the pens for one week after which the exploration tests were repeated in the same way as described before with all experimental groups on the farms.

As an additional parameter (total occupation with object), the number of animals occupied with the objects was determined as sum of animals sniffing, licking or chewing, horning or rubbing and jostling. For all behavioural parameters, the percentages of animals (i.e. in relation to the number of animals in the group) were calculated at all sample points. For the groups with a restricted animal-feeding place-ratio, also the percentages per object were calculated for all behaviour patterns except for lying.

For the analysis, mean percentages for the whole test hour were calculated for all parameters and, additionally, means for total occupation with object, feeding and lying over the first, second, third and fourth quarter of the test hour.

For statistical analysis of possible differences between barren (SF) and enriched (DL) pens, as well as between first test and repetition with or without simple enrichment of barren pens (SF), linear mixed effects models were used (Pinheiro and Bates, 2000) and calculated using R 2.3.1 (<http://www.r-project.org>).

For comparison between DL and SF pens, the fixed effects were α_l, housing system (factor with two levels: DL or SF), β_m, test session (factor with 2 levels: 1st or 2nd test session), and $\alpha_l{:}\beta_m$, the interaction of these effects. Nested random effects were b_i, country, b_{ij}, farm and b_{ijk}, test session. This resulted in a model of the form:

$$y_{ijklm} = \mu + b_i + b_{ij} + b_{ijk} + \alpha_l + \beta_m + \alpha_l{:}\beta_m + \varepsilon_{ijklm}$$

Effects of the simple enrichment of SF pens were analysed with the model:

$$y_{ijkl} = \mu + b_i + b_{ij} + \alpha_k + \beta_l + \alpha_k{:}\beta_l + \varepsilon_{ijkl}$$

with the intercept μ, the fixed effects α_k, presence of simple enrichment (factor with 2 levels: yes or no), β_l, test repetition (factor with 2 levels: 1st test or repetition of the test) and $\alpha_k{:}\beta_l$, the interaction of the fixed effects and the nested random effects b_i, farm and b_{ij}, group.

Statistical assumptions in these models are that $\varepsilon_{ijkl(mno)} \sim N(0, \sigma^2)$ iid, $b_i \sim N(0, \sigma1^2)$ iid, $b_{ij} \sim N(0, \sigma_2{}^2)$ iid and $b_{ijk} \sim N(0, \sigma32)$ iid (iid= independently identically distributed). These assumptions, homoscedasticity and independence of the residuals from the explanatory variables were checked using graphical analysis of residuals. The response variables for the comparison of the housing systems (DL and SF) were arcsin square root transformed.

18.4.2 Results

When calculating the percentages per animal, licking or chewing, horning or rubbing as well as the total occupation with object was more frequent in SF than in DL ($F_{1,18}$=8.877, p=0.008; $F_{1,18}$=5.882, p=0.026; $F_{1,18}$=12.968, p=0.002). No significant differences between bulls in SF and DL pens could be detected with regard to the behaviour patterns near to object without occupation and sniffing (Table 18.2). Jostling was observed only rarely, however, there was a significant interaction of housing and test session, with the bulls in DL showing more jostling in the second session of the day compared to the first and with the bulls in SF vice versa (housing system-test session-interaction, per object: $F_{1,16}$=7.801, p=0.013; per animal: $F_{1,16}$=6.084, p=0.0253).

Feeding was less frequent in the second test session (per object: $F_{1,16}$=20.187, p=0.0004; per animal: $F_{1,16}$=24.129, p=0.0002) and this difference was more pronounced in the bulls kept in SF pens (housing system-test session-interaction; per object: $F_{1,16}$=4.682, p=0.046; per animal: $F_{1,16}$=6.779, p= 0.0192). The percentage of lying animals was higher in the second test session than in the first ($F_{1,15}$=15.61, p=0.0013) and this difference was greater for the bulls in SF than for bulls in DL (housing system-test session-interaction; $F_{1,15}$=9.87, p= 0.0067).

When looking at the single quarters of the test hour, total occupation with object decreased over time with markedly higher values in the first 15 minutes and only slight differences between the third and fourth quarter of the test (Table 18.3). This was true for both test sessions. Total occupation with object was more frequent in the SF pens than in the DL pens when percentages were calculated per animal ($F_{1,18}$=9.208, p=0.0071; $F_{1,18}$=9.412, p=0.0066; $F_{1,18}$=5.34, p=0.0329; $F_{1,18}$=7.127, p=0.0156) but no difference could be seen in the percentages calculated per object. If calculated per object presented, an interaction of housing system and test session could be detected in the first quarter of the test hour with the animals in DL exploring the objects more in the second test session, whereas the animals in SF did not show such a difference ($F_{1,16}$=5.731, p=0.0293). In all quarters of the test hour, the animals fed on average less in the second test session (per object: $F_{1,16}$=7.514, p=0.0145; $F_{1,16}$=7.755, p=0.0133; $F_{1,16}$=19.86, p=0.0004; $F_{1,16}$=7.592, p=0.0141; per animal: $F_{1,16}$=7.321, p=0.0156; $F_{1,16}$=8.29, p=0.0109; $F_{1,16}$=22.912, p=0.0002; $F_{1,16}$=9.651, p=0.0068), though, this difference was only slight for the DL bulls in the third quarter (housing system-test session-interaction; per object: $F_{1,16}$=11.069, p= 0.0043; per animal: $F_{1,16}$=12.552, p=0.0027). Lying increased during the test hour. In general, more animals lay in the second test session ($F_{1,15}$=20.977, p=0.0004; $F_{1,15}$=7.48, p=0.0153; $F_{1,15}$=5.545, p=0.0326; $F_{1,16}$=15.216, p=0.0013). Though, in the second and third quarter, this pattern was reversed for the DL animals (housing system-test session-interaction; $F_{1,15}$=7.51, p=0.0152; $F_{1,15}$=10.244, p=0.006). At the end of the test hour, the percentage of lying animals was higher in the DL pens than in the SF pens ($F_{1,18}$=4.515, p=0.0477).

TABLE 18.2 Percentage (means; minimum and maximum values in parentheses) of animals performing different behaviour patterns in the exploration test (1h means) depending on housing system (deep litter, DF and slattted floors, SF) and test session on an experimental day.

Behaviour pattern	Housing system				Statistics[a]	
	DL		SF			
	Test session 1	Test session 2	Test session 1	Test session 2	Per object	Per animal
Feeding	.23 (.08–.78)	.19 (.03–.58)	.23 (.05–.55)	.12 (.03–.29)	S, I	*S, I*
	.16 (.05–.26)	*.12 (.03–.28)*				
Near to object without occupation	.07 (.04–.15)	.08 (.02–.14)	.09 (.01–.17)	.08 (0–.27)	ns	*ns*
	.06 (.02–.15)	*.06 (.02–.14)*				
Sniffing	.03 (0–.09)	.04 (0–.09)	.04 (0–.10)	.04 (0–.11)	ns	*ns*
	.02 (0–.06)	*.03 (0–.05)*				
Licking or chewing	.05 (0–.12)	.06 (.02–.13)	.07 (.03–.13)	.07 (.03–.16)	(H)	*H*
	.05 (0–.12)	*.05 (.02–.13)*				
Jostling	.01 (0–.03)	.02 (.01–.03)	.02 (0–.05)	.01 (0–.04)	I	*I*
	.01 (0–.02)	*.01 (0–.03)*				
Horning or rubbing	.06 (.01–.09)	.07 (.02–.15)	.06 (.01–.15)	.07 (.02–.20)	ns	*H*
	.05 (0–.09)	*.06 (.01–.15)*				
Total occupation with object	.14 (.03–.29)	.19 (.07–.28)	.19 (.13–.25)	.18 (.09–.28)	(H), (I)	*H*
	.13 (.01–.22)	*.15 (.06–.28)*				
Lying	*.25 (.04–.50)*	*.26 (.06–.51)*	*.07 (0–.27)*	*.32 (0–.76)*		*(H), S, I*

Notes: [a] significant results (p<.05, in parenthesis if p<.1) for housing system (H), test session (S), interaction (I) and no significance (ns); values calculated per object and, in italics, per animal.

TABLE 18.3 Percentage (means; min. and max. values in parentheses) of animals performing different behaviour (means for quarter I–IV of test session) depending on housing system (deep litter, DF and slattted floors, SF) and test session.

Quarter of test session	Behaviour pattern	Housing system				Statistics[a]	
		DL		SF			
		Test session 1	Test session 2	Test session 1	Test session 2	Per object	Per animal
I.	Feeding	.19 (.01–.68)	.10 (0–.50)	.16 (0–.54)	.11 (.01–.26)	S	*S*
		.13 (.01–.24)	*.06 (0–.19)*				
	Total occupation with object	.23 (.03–.51)	.35 (.05–.60)	.33 (.16–.49)	.31 (.15–.52)	I	*H, (I)*
		.20 (.01–.38)	*.27 (.05–.60)*				
	Lying	*.14 (0–.47)*	*.25 (0–.57)*	*.03 (0–.14)*	*.25 (0–.58)*		*S*
II.	Feeding	.29 (.11–.90)	.25 (.01–.61)	.26 (0–.90)	.16 (.02–.47)	S	*S*
		.21 (.10–.38)	*.16 (.01–.35)*				
	Total occupation with object	.13 (0–.28)	.18 (.06–.31)	.18 (.09–.39)	.18 (.09–.30)	ns	*H*
		.12 (0–.21)	*.15 (.04–.31)*				
	Lying	*.23 (0–.58)*	*.19 (0–.48)*	*.04 (0–.20)*	*.26 (0–.65)*		*S, I*
III.	Feeding	.23 (.04–.78)	.23 (.01–.57)	.28 (.07–.77)	.11 (0–.33)	S, I	*S, I*
		.16 (.04–.28)	*.15 (.01–.33)*				
	Total occupation with object	.12 (.03–.21)	.12 (.01–.18)	.14 (.03–.29)	.13 (0–.28)	ns	*H*
		.10 (.01–.21)	*.09 (.01–.18)*				
	Lying	*.32 (0–.52)*	*.22 (0–.47)*	*.08 (0–.41)*	*.34 (0–1.00)*		*S, I*
IV.	Feeding	.19 (.05–.78)	.20 (0–.66)	.22 (.05–.38)	.09 (0–.37)	S, (I)	S, (I)
		.13 (.03–.29)	*.12 (0–.35)*				
	Total occupation with object	.09 (.03–.19)	.09 (0–.19)	.11 (.03–.16)	.12 (.01–.27)		H
		.08 (.01–.16)					
	Lying	*.34 (.02–.74)*	*.49 (0–.99)*	*.12 (0–.50)*	*.44 (0–.83)*		*H, S, (I)*

Notes: [a] significant results (p<.05, in parenthesis if p<.1) for housing system (H), test session (S), interaction (I) and no significance (ns); values calculated per object and, in italics, per animal.

No effects of the simple enrichment over one week in SF pens on the behaviour patterns recorded could be detected (Tables 18.4 and 18.5). Independently of the treatment, more feeding ($F_{1,20}$=4.63, p=0.0438) and less lying ($F_{1,20}$=6.438, p=0.0196) were observed during the repeated test. The animals showed also slightly more licking or chewing ($F_{1,20}$=7.116, p=0.0148) but less horning or rubbing ($F_{1,20}$=16.59, p=0.0006). Taking into

TABLE 18.4 Percentage (means; minimum and maximum values in parentheses) of animals performing different behaviour patterns in the exploration test (1h means) depending on enrichment conditions and repetition of the test.

Behavior pattern	Enriched		Non enriched		Statistics[a]
	1st test	Repetition	1st test	Repetition	
Feeding	.20 (.04–.47)	.24 (.11–.36)	.19 (.03–.55)	.22 (.03–.49)	R
Near to object without occupation	.12 (.03–.27)	.13 (.07–.17)	.10 (.03–.20)	.10 (.04–.20)	(E)
Sniffing	.06 (.02–.11)	.06 (.02–.13)	.06 (.02–.10)	.06 (.03–.12)	ns
Licking or chewing	.07 (.03–.16)	.08 (.01–.20)	.06 (.03–.10)	.09 (.02–.15)	R, (I)
Jostling	.01 (0–.02)	.01 (0–.02)	.01 (0–.02)	.01 (0–.02)	ns
Horning or rubbing	.04 (.01–.06)	.03 (.01–.05)	.04 (.01–.08)	.02 (.01–.04)	R
Total occupation with object	.18 (.12–.26)	.17 (.06–.30)	.17 (.09–.24)	.18 (.08–.25)	ns
Lying	.12 (0–.41)	.06 (0–.31)	.25 (.01–.76)	.11 (0–.41)	R

Notes: [a] significant results (p<0.05, in parenthesis if p<0.1) for enrichment (E), repetition (R), enrichment-repetition-interaction (I) and no significance (ns).

TABLE 18.5 Percentage (means; minimum and maximum values in parentheses) of animals performing different behaviour patterns in the exploration test (means for quarter I-IV of test session) depending on enrichment conditions and repetition of the test.

	Behavior pattern	Enriched		Non-enriched		Statistics[a]
		1st test	Repetition	1st test	Repetition	
I.	Feeding	.18 (.01–.54)	.22 (.01–.44)	.13 (0–.25)	.15 (.01–.44)	ns
	Total occupation with object	.30 (.15–.45)	.29 (.10–.45)	.33 (.18–.52)	.32 (.14–.44)	ns
	Lying	.07 (0–.48)	.05 (0–.25)	.16 (0–.58)	.06 (0–.20)	ns
II.	Feeding	.22 (0–.54)	.27 (.15–.46)	.27 (.03–.90)	.28 (.08–.45)	ns
	Total occupation with object	.18 (.09–.28)	.15 (.04–.29)	.15 (.09–.24)	.16 (.03–.25)	ns
	Lying	.07 (0–.31)	.03 (0–.33)	.19 (0–.65)	.05 (0–.20)	R
III.	Feeding	.24 (.06–.47)	.24 (.03–.42)	.20 (0–.77)	.28 (.03–.51)	ns
	Total occupation with object	.13 (.03–.21)	.11 (.03–.22)	.10 (0–.20)	.12 (.07–.21)	ns
	Lying	.12 (0–.41)	.03 (0–.31)	.30 (0–1.00)	.11 (0–.41)	(E), R
IV.	Feeding	.18 (.03–.37)	.23 (.02–.47)	.18 (0–.38)	.17 (.01–.63)	ns
	Total occupation with object	.12 (.08–.16)	.11 (.03–.31)	.09 (.01–.16)	.12 (.01–.23)	ns
	Lying	.22 (0–.56)	.13 (0–.36)	.34 (0–.80)	.22 (0–.82)	R

Notes: [a] significant results (p<0.05, in parenthesis if p<0.1) for enrichment (E), repetition (R) and no significance (ns).

account the single quarters of the test hour, the decrease in lying during the repeated test was only due to a significant decrease in the second, third and fourth quarter ($F_{1,20}$=6.865, p=0.0164; $F_{1,20}$=7.127, p=0.0147; $F_{1,20}$=4.598, p=0.0445).

18.5 DISCUSSION

Although from a practical point of view it was our impression that we succeeded in developing an on-farm test for exploratory behaviour in beef bulls, we did not manage to validate this test as indicator of daily levels of exploratory behaviour. Though bulls kept in barren housing conditions (fully slatted floor pens) explored the novel object more, differences were only slight and values of groups varied a lot within housing systems and ranges of the different housing conditions overlapped.

One of the weaknesses of our approach was the lack of knowledge about the actual levels of exploratory behaviour under the various conditions. In fact, also deep litter pens still may provide a rather monotonous environment, and the 'enrichment' over one week was very simple indeed. However, long-term observations of the daily behaviour in 61 groups of bulls would not have been feasible within the frame of this study. The housing systems that we have selected for this study, those with deep litter and fully slatted floors, already represent extremes of the systems typically found in intensive beef fattening. Even though the inclusion of a more extensive system (e.g. bulls on pasture) would probably have provided a better basis for interpretation of these results, the fact remains that the test, as developed, is not sensitive enough to detect possible differences between usual commercial systems. Another possible reason for the lack of differences is that ruminants might have a relatively low propensity of inquisitive exploration compared to other species such as pigs or rats (Wood-Gush and Vestergaard, 1989). Inquisitive exploration according to Berlyne (1960; c.f. Wood-Gush and Vestergaard, 1989) is the attempt to make a change rather than respond to a change whereas inspective exploration is the inspection of a particular object. Both are classified as types of intrinsic exploration. From own observations we would classify cattle clearly as neophilic, i.e. highly motivated to perform inspective and inquisitive exploration (e.g. Emmerig, 2004). Another aspect worthwhile considering is that the response of barrenly kept animals to a novel object could be much lower than expected if apathy due to severe deprivation of intrinsic exploration has developed (Wood-Gush and Vestergaard, 1989; Wemelsfelder and Birke, 1997). Possible species-specific differences in exploratory behaviour and the development of exploration during life in different environments are questions certainly deserving more investigation in the future.

One confounding factor on the German farms was the difference in feeding place to animal ratio between housing systems, with most of the deep litter pens having less than one feeding place per animal. Thus, if the occupation per animal would have been the same between SF and DL pens, necessarily occupation per object would have been as many times longer as there were animals per object. On the other hand, when looking at time of occupation with an object per animal, exploration will likely be underestimated because the potential for interactions with the object were smaller per animal in the DL compared to the SF pens. These values would only realistically reflect the individual motivation to explore if no social effects such as social facilitation or inhibition were in place. Therefore the truth for DL bulls will lie somewhere between the two extremes of the values per object and per animal. In this study, even in the groups with less than one feeding place and thus one object per animal, however, 'free' objects were always available (data not shown). Therefore, always more animals could have had access to the objects. This may indicate that the calculation on a per animal basis does not underestimate the true incidence of explorative behaviours.

Regarding the interactions between housing systems and test session, it is likely that again the number of feeding places per animal has played a role. The pens with a feeding place to animal ratio of 1:1 obviously allowed more synchronous feeding behaviour with a more pronounced resting phase already two hours after feeding, whereas in the pens with less feeding places there were still active animals at the feeding gate. In the first hour, competition for food was probably reducing interest in the novel objects in the DL bulls while possibly competition for lying space in SF bulls had an interest reducing effect in the second hour. This would not indicate a high enough attractiveness of the novel objects to profoundly influence the usual activity patterns, and makes the test even more problematic for application under different housing conditions.

Because of the difficulty to take into account possible effects of different animal to feeding place ratios, later a large proportion of the data from this report have since been complemented with further data in order to achieve nearly similar animal to feed place ratios (Schulze Westerath et al., 2009). Significant differences between DL and SF bulls remained for total occupation with objects as well as licking or chewing. Also, lying increased more in the second test session in SF bulls compared to DL bulls, but without effects on exploration levels.

Although it appeared that the objects did not elicit fear (latency to approach objects equal to 0 seconds on all farms), the increased feeding and decreased lying in the repeated tests could suggest that fear might have kept bulls away from the feeding table in the first test. Also the slight change in proportions of the different behaviours at the objects may indicate some form of habituation. However, from the data available this can only be speculative.

18.6 CONCLUSIONS

We did not succeed in developing a valid on-farm test of exploratory behaviour for beef bulls. Although bulls kept on fully slatted floors (assumed low daily stimulation) showed more exploration of novel objects than those kept in littered pens (assumed higher stimulation), differences were only slight. Moreover, no differences could be detected after a simple, short-term stimulation under barren conditions.

19

RELIABILITY OF MEASURES OF SOCIO-POSITIVE AND PLAY BEHAVIOUR IN DAIRY AND BEEF CATTLE

H. Schulze Westerath, N. Brörkens, S. Laister, N. MacKintosh, C. Winckler and U. Knierim

19.1 SUMMARY

During recent years the interest in indicators of good welfare has increased. In dairy and beef cattle, possible positive behavioural indicators are social licking, horning and playing. Social licking is an affiliative behaviour and has been attributed tension-reducing and calming effects. It might therefore also be an indirect indicator of the social stress level in a herd and make the interpretation of incidences on farm level more difficult. Social and non-social play behaviour has often been suggested as an indicator of good welfare state since particularly young animals are only motivated to play if their primary needs are satisfied. Horning has been associated with positive indicators as it often occurs together with running games or playful mounting and has been described in the context of mock fighting.

It was the aim of this study to investigate social licking, horning and locomotor/object play as on-farm measures in dairy cows and fattening bulls with regard to feasibility, inter-observer reliability and short- to long-term intra-farm variability (consistency). For this purpose, continuous behaviour sampling was carried out on three days for 4 h each on 43 dairy (31 deep litter, 12 tie stalls) and 20 beef farms (10 deep litter, 10 fully slatted floors) in Austria, Germany and Italy. In beef farms, three weight classes were defined (I: 200-350 kg, M: 350-500 kg, F: >550 kg). Farm visits took place at approximately 60 and 180 days (tie stalls: 120 days) after the first visit. Inter-observer reliability was tested using direct observations (6 dairy, 2 beef farms) and from video clips (n=55).

Play behaviour was only very rarely observed and is therefore regarded to be not useful for short-term on-farm recordings in fattening and adult cattle. This is also the case for horning in dairy cows. Inter-observer reliability (IOR) for social licking was high in both

direct observations (3 observers: Kendall's W=0.96) and observations from videos (W=0.93). IOR for horning was lower during live observations (W=0.70). This was probably due to misinterpretation of the definition, as it increased to good levels in the video clip observations (W=0.86) after the definition had been refined.

Intra-farm consistency was highest for horning in fattening bulls (W=0.82-0.86). For social licking it was rather low for the youngest class (W=0.64) but increased to acceptable levels when the two heavier weight classes were merged (W=0.75). In dairy cattle, there was substantial intra-farm variation across time for social licking (W=0.57/loose housing and 0.74/tie stalls) and horning (W=0.68/loose housing and 0.48/tie stalls).

In conclusion, for on-farm welfare assessment in fattening bulls the inclusion of social licking and horning is recommended; medium weight and finishing bulls should not be distinguished. For dairy cattle, for none of the behaviours assessed it proved to be possible to reliably record them in single short-term observations.

19.2 INTRODUCTION

Whereas most approaches to welfare assessment are based on indicators of impaired welfare, there is an increasing interest in indicators of good welfare. Broom (1999) stresses that behavioural indicators of pleasure provide measures for good animal welfare. Environmental control, play behaviour and positive social relations may be considered as main components of good welfare (Fagen, 1981; Lawrence, 1987). 'Positive' indicators to our knowledge have not been applied in on-farm welfare monitoring systems so far.

Social licking is regarded important for building affiliative bonds and reducing tension or anxiety between group members (Bouissou et al., 2005; Sambraus, 1969; Sato et al., 1993). Wood (1977) even suggests social licking as a parameter to describe an affiliative scale of social distance within cow herds as an alternative to scales of social distance based on agonistic interactions. Short-term calming effects in terms of a reduction in heart rate have been demonstrated in primates (Boccia et al., 1989; Aureli et al., 1999) and cattle (Sato et al., 1993). However, it seems debatable whether high levels of social licking activity always indicate a better welfare state at herd level when compared with herds which show less social licking.

Social and non-social play behaviour have often been suggested as an indicator of good welfare state since particularly young animals are only motivated to play if their primary needs are satisfied (Fagen, 1981; Lawrence, 1987). In calves, play is mainly expressed as locomotor and social activities. In addition, Jensen et al. (1998) described play activities directed towards the environment (object and straw butting and rubbing). However, playing is only rarely observed in adult animals. Horning has been described in the context of mock fighting and has been associated with positive indicators as it often occurs together with running games or playful mounting (Reinhardt & Reinhardt, 1982).

The objective of this study was to investigate social licking, horning and locomotor/object play as on-farm measures in dairy cows and fattening bulls with regard to feasibility, inter-observer reliability and short- to long-term intra-farm variability (consistency). The latter addressed the question how representative single recordings are with regard to possible changes due to seasonal effects and/or related to management (e.g. regrouping, purchase of new animals). This is less relevant when welfare assessment systems are used as decision support tools (Sørensen et al., 2001). However, if welfare assessment protocols are going to be used for certification purposes with infrequent or even single assessments only, the representativeness of recordings with regard to the longer-term situation on the farm becomes especially important.

19.3 METHODS

19.3.1 Inter-observer Reliability Testing (IORT)

IORT of positive indicators was carried out with (1) direct observations on six dairy and two beef farms and (2) analysis of video clips. For the investigation of inter-observer reliability no distinction was made between dairy cows and beef bulls as behaviour patterns are the same in both categories of cattle.

On-farm IORT

On-farm IORT was carried out three times at different stages of the project with different numbers of observers being present at each date. The first took place after two days of initial training which included discussions, video and on-farm training (1 deep litter dairy farm). The second IOR meeting was about 50 days after observers A and B had started on-farm data collection; two additional observers (C and D) participated, who had received about half a day of training. The final on-farm IOR testing took place after finishing data collection in order to check whether inter-observer agreement had changed (see Table 19.1).

Observers were always located near to each other in the barn allowing a free view on the area observed.

TABLE 19.1 Overview of on-farm IOR meetings.

Meeting	mm-yyyy	Number of farms visited	Sample size (pens, segments)			
			A	B	C	D
1 Germany	07-2005	4	35	35	–	–
2 Austria	09-2005	2	10	10	7	10
3 Italy	06-2006	2	15	15	15	–
	Total	8	60	60	22	10
	Dairy	6	50	50	22	10
	Beef	2	10	10	–	–

Spearman rank correlation (r_s) was used to test agreement within pairs of observers. Additionally Kendall's coefficient of concordance (W) was calculated for agreement between 3 and 4 observers, respectively.

IORT Using Videos

After completion of the on-farm data collection and after the final on-farm IOR meeting in June 2006, three trained observers (A, B & C) separately watched 55 video sequences of each about 2 to 14 minutes (in total 07:05 hours). The video observations followed the same rules as provided in the instructions for data collection.

Again Spearman rank correlations and Kendall's coefficient of concordance (W) were used based on the frequencies for each behaviour recorded from the video clips.

19.3.2 Investigation of Intra-farm Consistency

In total, 43 dairy farms (20 cubicle, 11 deep litter, 12 tie stall systems; herd size 12 – 150 cows (Table 19.2)) and 20 beef fattening farms (10 deep litter, 10 fully slatted floor systems; animals per farm: 30-220, 5-27 bulls per pen) in Austria, Germany and Italy (only dairy) were included in the study. The dairy cows belonged to different breeds with Holstein Friesian, Simmental-Fleckvieh and Brown Swiss being the most prevalent breeds. The fattening bulls were Simmental-Fleckvieh (S-FV), Limousin and S-FVxLimousin crosses. Other breeds such as Belgian Blue, Brown Swiss, Holstein Friesian, Grauvieh or Charolais were also kept in small numbers.

Observations of 'positive' behaviours as given in Table 19.3 were carried out on three days at intervals of approximately 60 and 180 days after the first visit (beef cattle farms, dairy loose housing systems). In dairy farms with tie stalls, the third visit took place 120 days after the first visit (Figure 19.1).

TABLE 19.2 Overview of dairy farms visited in each country

	Austria	Germany	Italy	Total
Cubicles	8	8	4	20
Deep litter	3	4	4	11
Tie stalls	6	6	–	12
Total	17	18	8	43

TABLE 19.3 List of behaviours observed and definitions.

Behaviour	Definition
Social Licking (SL)	The actor touches with its tongue any part of the body (head, neck, torso, legs, tail) of another group mate except for the anal region or the prepuce. If the actor stops licking for more than 10 s and then starts licking the same receiver again, this is recorded as a new bout. It is also taken as a new bout, if the actor starts licking another receiver or if there's a role reversal between actor and receiver.
Horning (HO)	Head play with physical contact of two animals: The animals are rubbing their foreheads, horn bases or horns against the head or neck of one another without obvious agonistic intention. None of the opponents takes advantage of the situation in order to become a victor (Reinhardt and Reinhardt 1982). It is taken as a new bout if the same animals start horning after 10 seconds or more or if the horning partner changes.
Playing (PL)	1.) Locomotor play: Galloping (running very fast, often with tail lifted up high) and Bucking (jumping in the air with all four feet or hind feet off the ground, often with tail lifted up high). 2.) Playing with equipment: Rubbing horns or horn base against any equipment in the barn. However, the regular rubbing on equipment with the intention to scratch is not taken as playing.

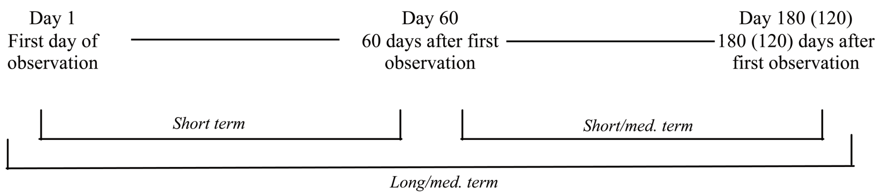

FIGURE 19.1 Observation scheme of the study for dairy cows (tie stall herds in parenthesis) and fattening bulls.

Data Recording on Dairy Farms

On each farm the lactating dairy cows were observed excluding separate groups of heifers, dry or periparturient cows, as well as cows in hospital pens. Behaviour performed by or with animals in heat were not recorded.

Behaviour was recorded using continuous behaviour sampling (Table 19.3) and observations lasted for 4 h after morning feeding (or after feed had been pushed up). The observer was positioned on the feed bunk on an elevated observation chair.

In herds larger than 25 cows, the observations were carried out in segments of the barn which were expected to contain on average not more than 25 cows per segment. These segments covered all accessible areas (lying areas, feeding places, concentrate feeders, outdoor loafing areas etc.). The duration of continuous observations within each segment was adjusted to the number of segments so that each segment was observed at least once per two hours (minimum observation period 10 min). Each segment was observed at least twice.

The number of animals which were feeding, standing/walking and lying within the segment was recorded at the beginning and at the end of each observation period. Data were then analysed as the mean number of events per cow and hour, taking the absolute frequency of behaviours, the duration of observations per pen/segment and the average number of animals in the pen/segment during the observation into account.

From all values obtained on segment level the mean incidence at herd level was calculated (occurrence of behaviour/animal*hour).

Spearman rank correlations and Kendall's coefficient of concordance were used to test repeatability between visits (1-60-180/120).

Data Recording on Beef Farms

Three weight classes were defined in line with the literature and common farming practice:

- initial fattening period (I): (200–350 kg)
- medium fattening period (M): (350–550 kg)
- finishing fattening period (F) (≥550 kg)

Pens holding less than three animals were excluded from observations. Behaviour was recorded using continuous behaviour sampling and observations lasted 4 h after morning feeding (or after feed had been pushed up). The observer was positioned on the feed bunk on an elevated observation chair.

All weight classes present were observed for equal periods within each observation hour and each pen was observed at least twice during the 4 hour period. Pens with more than 25 bulls were divided into two or more segments (see dairy cattle).

Data processing within each weight class followed the same rules as for dairy cattle.

19.3.3 Decision on Measures

We followed a stepwise approach in deciding on the usefulness of the behavioural measures for on-farm welfare assessment protocols:

1. On-farm incidence had to exceed 0.10 events/animal*hour to allow for reliable and feasible recording as well as differentiation between farms;
2. The remaining parameters were then checked for inter-observer reliability (r_s/Kendall's W≥0.70).
3. Intra-farm consistency was used for the final evaluation only using Kendall's W of 0.70 as threshold.

19.4 RESULTS

19.4.1 Incidences of Positive Indicators

Social licking was the most frequent behaviour in both dairy and beef cattle (Tables 19.4 and 19.5). In dairy tie stalls, the average incidence was about four times higher than in loose housing systems (median 0.98/animal*hour vs. 0.25/animal*hour). In beef bulls it was shown most frequently during the initial fattening period and least in finishing bulls.

Table 19.4 Descriptive measures of frequencies of positive indicators in dairy cows (events/animal*hour; n = farms).

Behaviour	Housing system	Day	Median	Mean	Min	Max	SD	Var	N
Social licking	Loose housing	1	.32	.37	.05	1.11	.21	.04	31
		60	.21	.28	.02	.73	.20	.04	31
		180	.20	.28	.00	2.13	.38	.15	31
		overall	.25	.31	.00	2.13	.28	.08	93
	Tie stalls	1	1.25	1.26	.17	2.40	.79	.62	12
		60	.87	1.09	.34	2.60	.72	.52	12
		120	.98	.96	.42	1.61	.43	.18	12
		overall	.98	1.10	.17	2.60	.66	.43	36
Horning	Loose housing	1	.04	.08	.00	.32	.09	.01	31
		60	.06	.09	.00	.33	.09	.01	31
		180	.07	.12	.00	.68	.16	.02	31
		overall	.06	.10	.00	.68	.12	.01	93
	Tie stalls	1	.06	.11	.00	.45	.14	.02	12
		60	.00	.03	.00	.15	.04	.00	12
		120	.04	.07	.00	.21	.08	.01	12
		overall	.04	.07	.00	.45	.10	.01	36
Playing	Loose housing	1	.00	.01	.00	.05	.01	.00	31
		60	.00	.01	.00	.13	.02	.00	31
		180	.00	.00	.00	.08	.02	.00	31
		overall	.00	.01	.00	.13	.02	.00	93
	Tie stalls	1	.00	.00	.00	.00	.00	.00	12
		60	.00	.00	.00	.00	.00	.00	12
		120	.00	.00	.00	.00	.00	.00	12
		overall	.00	.00	.00	.00	.00	.00	36

TABLE 19.5 Descriptive measures of frequencies of positive indicators in fattening bulls (events/animal*hour; n = farm visits).

Behaviour	Weight class	Median	Mean	Min	Max	SD	Var	N
Social licking	I	.85	.93	.00	3.71	.80	.64	56
	M	.73	.94	.10	2.98	.71	.50	57
	F	.68	.95	.00	3.83	.87	.75	57
Horning	I	.40	.70	.00	2.81	.80	.63	56
	M	.68	1.02	.00	3.18	.95	.90	57
	F	.53	1.04	.00	3.71	1.02	1.04	57
Playing	I	.00	.07	.00	1.04	.16	.03	56
	M	.00	.04	.00	.29	.07	.00	57
	F	.00	.05	.00	.38	.10	.01	57

Compared to social licking horning was much less frequent in dairy cows, whereas it was almost as frequent as social licking in fattening bulls (overall median 0.52/animal*hour).

Playing occurred only very rarely with the maximum frequency shown in initial fattening bulls.

19.4.2 INTER-OBSERVER RELIABILITY

On-farm Observations

The best inter-observer reliability (IOR) was found for social licking with correlation coefficients above 0.90 for almost all pairs of observers as well as for the coefficients of concordance (Table 19.6). The agreement between observer A and B furthermore improved with the number of sessions.

For horning the agreement between observers was lower and reached satisfactory levels ($r_s > 0.70$) only for some pairs of observers during the first two meetings.

Play behaviour could not be tested for inter-observer reliability because it occurred only once during the on-farm observations.

Video observations

The total numbers of behaviours observed in 55 video clips is given in Table 19.7. Playing was observed twice by each observer, however in 4 different videos. Therefore playing again was omitted from further analysis.

As for live observations, from videos the highest inter-observer agreement was found for social licking (Kendall's W = 0.93; Table 19.8). IOR for horning (W=0.86) was slightly lower, but better than in on-farm observations due to improvements of the definition.

TABLE 19.6 Inter-observer reliability in different test sessions (Spearman rank correlation coefficients, Kendall's coefficient of condordance) for social licking and horning.

Behaviour	Meeting		Observer pairs						Kendall's W	
			A_B	A_C	A_D	B_C	B_D	C_D	ABC	ABCD
Social Licking	1	r_s	.91							
		p	.00							
		n	35							
	2	r_s	.98	.96	.98	1.00	1.00	1.00		
		p	.000	<.001	.000	.000	.000	.000		
		n	10	7	10	7	10	7		
	3	r_s	1.00	.82		.82				
		p	.000	<.001		<.001				
		n	15	15		15				
	Overall	r_s	.94	.91	.98	.91	1.00	1.00	.96	.99
		p	.000	.000	.000	.000	.000	.000	.000	.000
		n	60	22	10	22	10	7	22	7
Horning	1	r_s	.73							
		p	.000							
		n	35							
	2	r_s	.76	.84	.39	.50	.22	.38		
		p	.010	.019	.270	.253	.536	.394		
		n	10	7	10	7	10	7		
	3	r_s	.21	.53		.53				
		p	.450	.040		.044				
		n	15	15		15				
	Overall	r_s	.59	.69	.39	.56	.22	.38	.70	.59
		p	.000	<.001	.270	.007	.536	.394	.000	.010
		n	60	22	10	22	10	7	22	7

Notes: A, B, C = observers.

TABLE 19.7 Number of events observed by observer A, B and C scored from 55 video clips.

Behaviour	Number of observed behaviours		
	A	B	C
Social licking	55	51	58
Horning	32	39	37
Playing	2	2	2

TABLE 19.8 Spearman rank correlation coefficient (r_s) and Kendall's coefficient of concordance (W) for video inter-observer agreement (N=55).

Behaviour		Observer pairs			Kendall's W
		A_B	A_C	B_C	
Social Licking	r_s	.89	.86	.92	.93
	p	<.001	<.001	<.001	<.001
Horning	r_s	.84	.81	.72	.86
	p	<.001	<.001	<.001	<.001

19.4.3 Consistency over Different Periods of Time (Intra-farm Variability)

In dairy cattle, the correlations between the three observation days (1–60–120/180) were inconsistent for social licking and horning in loose housed dairy herds and for horning in dairy cows kept in tie stalls (Table 19.9). Rather stable, but moderate correlations were found for social licking in tethered animals. Therefore, the overall consistency (Kendall's W) for social licking was higher in tie stalls than in loose housing systems. For horning it was lower and not significant in both housing systems.

In beef cattle, horning showed the highest short-, mid- and long-term as well as overall consistency with all correlation coefficients above 0.70 in weight classes I and F (all $p<0.01$) and Kendall's W above 0.80 in all weight classes (Table 19.10). Merging the data from weight classes M and F further improved consistency.

Social licking did not show a clear pattern with regard to short-, mid- or long-term consistency. Recordings at an interval of about 6 months in weight classes I and M were better correlated than on a short- or mid-term basis whereas the opposite was the case for finishing bulls. Merging data from medium weight and finishing bulls improved consistency measures to a limited extent (e.g. Kendall's W=0.75).

Reliability of Recordings from Reduced Observation Time

From the on-farm data, correlations between the first or the first two hours of observation time and the total observation time (4 h) were calculated in order to check for effects of reducing the time of observation.

In fattening bulls, for both social licking and horning the overall correlation coefficients between incidences derived from 4h-observations and 2h- or 1h-observations were equal or above 0.90 ($p<0.001$; Table 19.11). However, when the observation period was

Table 19.9 Spearman Rank correlation coefficients (r_s) for short- (1_60, 60_120), mid- (1_120, 60_180) and long-term (1_180) intra-farm consistency and Kendall's coefficients of concordance (W) for social licking and horning in dairy cattle.

Housing system	Behaviour		short-term 1_60	mid-term 60_180	long-term 1_180	Kendall's W
Loose housing	Social licking	r_s	.33	.57	.18	.57
		p	.070	<.001	.330	<.001
	Horning	r_s	.69	.54	.34	.68
		p	.000	.002	.059	.000
			short-term 1_60	short-term 60_120	mid-term 1_120	
Tie stalls	Social licking	r_s	.66	.65	.50	.74
		p	.020	.022	.095	<.001
	Horning	r_s	.38	.19	.10	.48
		p	.220	.562	.752	.113

TABLE 19.10 Spearman rank correlation coefficients (r_s) for short- (1_60), mid- (60_180) and long-term (1_180) intra-farm consistency and Kendall's coefficient of concordance (W) for social licking and horning in fattening bulls.

Behaviour	Weight class		short-term 1_60	mid-term 60_180	long-term 1_180	Kendall's W
Social licking	I	r_s	.29	.38	.71	.64
		p	.246	.113	<.001	.001
	M	r_s	.57	.54	.69	.73
		p	.014	.021	.001	.000
	F	r_s	.55	.36	.17	.55
		p	.018	.146	.475	.014
	M & F[1]	r_s	.74	.53	.64	.75
		p	<.001	.024	.003	.000
Horning	I	r_s	.82	.76	.80	.86
		p	<.000	<.001	.000	.000
	M	r_s	.64	.91	.68	.82
		p	.004	.000	.001	.000
	F	r_s	.79	.80	.72	.85
		p	<.001	.000	<.001	.000
	M & F[1]	r_s	.82	.87	.86	.90
		p	.000	.000	.0000	.000

Notes: [1] M & F = mean of medium and finishing weight class.

TABLE 19.11 Spearman rank correlation coefficients (r_s) for incidence of social licking and horning in loose housed dairy and beef cattle between 1 and 4 hours and 2 and 4 hours of observation.

	Behaviour	Day	1h_4h	p	2h_4h	p	n
Dairy (loose housed)	Social licking	1	.43	.015	.63	<.001	31
		60	.73	<.001	.84	<.001	31
		180	.56	.001	.68	<.001	31
		Overall	.63	<.001	.78	<.001	93
	Horning	1	.53	.002	.81	<.001	31
		60	.61	<.001	.81	<.001	31
		180	.65	<.001	.87	<.001	31
		Overall	.60	<.001	.84	<.001	93
Beef	Social licking	1	.94	<.001	.96	<.001	20
		60	.89	<.001	.98	<.001	18
		180	.90	<.001	.99	<.001	19
		Overall	.94	<.001	.98	<.001	57
	Horning	1	.82	<.001	.95	<.001	20
		60	.93	<.001	.96	<.001	18
		180	.97	<.001	.98	<.001	19
		Overall	.90	<.001	.97	<.001	57

shortened to one hour the correlation coefficients on a day basis (i.e. observation period) markedly decreased (down to 0.82). In dairy cows, however, correlation coefficients for 2h-observations already were below 0.80 for social licking and averaged 0.84 for horning. Reducing the observation time to 1 hour resulted in correlations of 0.63 and 0.60 for social licking and horning, respectively.

Intra-farm consistency across time was recalculated using only the first two hours of observation (Table 19.12, hour 1+2) or the third and fourth hour of observation (hour 3+4). For loose housed dairy herds, coefficients of concordance for the first two hours remained almost unchanged, but further decreased when only data from hour 3+4 were taken into account This was also the case for fattening bulls.

Recommended parameters are presented in Box 19.1.

19.5 DISCUSSION

Social licking, horning and play behaviour had initially been chosen as potential 'positive' indicators. Playing however, was only very rarely observed in both dairy and beef cattle (median 0 events/animal*hour). Very low incidences make data recording difficult and would require long-term continuous observations in order to get reliable information on

TABLE 19.12 Kendall's coefficient of concordance for social licking and horning in loose housed dairy herds and fattening bulls for the first and second two hours of observation over three observation days.

Behaviour				Total observation time (4 h)	Hour 1+2	Hour 3+4
Dairy (loose housed)	Social licking		W	.57	.56	.55
			p	<.001	.002	.002
	Horning		W	.68	.65	.51
			p	.000	.000	.010
		n		31	29	29
Beef	Social licking	I	W	.64	.57	.59
			p	.001	.008	.005
		M	W	.73	.70	.64
			p	.000	.000	.001
		F	W	.55	.55	.55
			p	.014	.016	.016
	Horning	I	W	.86	.80	.55
			p	.000	.000	.014
		M	W	.82	.79	.74
			p	.000	.000	.000
		F	W	.85	.81	.74
			p	.000	.000	.000
		n		18	18	18

BOX 19.1 Recommended parameters.

Dairy cattle:	None
Fattening bulls:	Frequency of social licking
	Frequency of horning
	(*merging of medium weight and finishing classes*)

this type of behaviour, which is not feasible. It was therefore not considered for inclusion into on-farm assessment protocols for dairy and beef cattle. The use of play behaviour as a welfare indicator is therefore probably limited to calves. In dairy cattle, horning was also rarely observed (median <0.10 events/animal*hour) and it is questionable whether it can be reliably recorded. It should therefore not be included in protocols for dairy.

The inter-observer agreement (IOR) for social licking was high in both live observations and when the behaviour was scored from videos. Rank correlation coefficients for single pairs of observers were always above 0.80 (as suggested as threshold for good IOR by Martin & Bateson, 1992) and overall coefficients of concordance between multiple observers above 0.90. With regard to horning, inter-observer reliability in live observations only reached acceptable levels for single pairs of observers during the first two sessions and IOR deteriorated during the third meeting. This was on the one hand due to lack of training of the inexperienced observers (B, C) in the second meeting which emphasizes the necessity for thorough training. The drop in IOR can be explained by a misinterpretation of the behaviour description which developed at a later stage. Horning is a manifold behaviour including behavioural elements from rubbing to soft head butts and slight thrusts (Reinhardt & Reinhardt, 1982), which make a clear and unambiguous definition difficult. Intensive training of observers is therefore necessary. This is why a refined definition of horning (as given in Table 3) was used for the final IOR testing from videos, which resulted in acceptable to good agreement regarding this behaviour with Kendall's W=0.86.

Social licking occurred only moderately consistently in farms with tie stalls (Kendall's W=0.74), and showed no regular pattern in loose housed herds. There was also no evidence that variability is lower within shorter periods of time. This is in agreement with earlier small-scale studies, which revealed a high inter-day variation in cubicle housed dairy cows on the basis of observations during three consecutive days (Winckler et al., 2002), as well as inconsistent data from bimonthly farm visits (Winckler et al.,2007). Such high variability may be caused by various factors such as introduction of new group members, animals in heat, ambient temperature or human-animal interactions, which cannot be controlled for within the context of on-farm welfare assessment recordings. Inclusion of social licking into assessment protocols would therefore mean that single recordings will only reflect the actual activity of a herd with rather unclear significance for the welfare state of the herd (see validity of the measure).

In fattening bulls, social licking was moderately consistent and reached satisfactory levels, when medium and finishing weight classes were merged. However, horning was shown to be the most consistent behaviour in all weight classes (Kendall's W=0.82-0.86). Merging the medium and finishing weight classes further reduced intra-farm variability. Pooling of these classes could furthermore improve feasibility of the recordings since live weight estimation appeared to be rather difficult in heavier bulls (about 550 kg) and groups were often heterogeneous. A distinction between younger bulls (up to 350 kg) and older ones should still be made since the younger animals are not sexually mature yet.

With regard to feasibility, shortening the on-farm observations from 4 h to 2 h would only cause a minor information loss for fattening bulls. When calculations were based on the first two hours, consistency was not affected by the reduction of observation duration. This indicates that behaviour directly after feeding is more consistent than at later times of the day.

19.6 CONCLUSIONS

In conclusion, with regard to inter-observer repeatability and intra-farm consistency we suggest to include social licking and horning for fattening bulls as positive indicators. None of the potential measures for dairy cattle proved to be reliable.

20

RELIABILITY OF MEASURES OF SOCIO-POSITIVE AND PLAY BEHAVIOUR IN CALVES

H. Leruste, J. Lensink and C.G. van Reenen

20.1 SUMMARY

Playing and social behaviours can have a positive impact on the welfare of calves. High levels of these behaviours may also result from a certain deficiency in the environment and, therefore, reflect a form of redirected behaviour or a behavioural rebound following deprivation. Measuring their incidence is an important point in determining levels of welfare. The aim of this study was to develop a valid method for on-farm assessment of the level of play and social behaviours (licking, mounting and fighting) in group-housed calves. This method should be simple, fast, feasible in on farm conditions, not biased by an observer effect and consistent in time.

Observations were performed on 20 farms in France. Calves were housed in large groups (between 20 and 64 calves per pen) and fed with an automatic milk dispenser (AMD). Two randomly chosen pens were observed per farm. Observations were performed simultaneously by two trained observers. The observations took place on 4 days: 2 observations at a two day interval at week 9 (after the arrival of the calves in the farm) and 2 observations at a two day interval at week 13.

Overall, the frequency of the four behaviours was low. Despite this, for all four measures an important farm effect was found ($p<0.001$) showing that these measures permit to distinguish between farms. Observers agreed on the level of plays ($r=0.95$; $p<0.001$; $n=20$), social licking ($r=0.90$; $p<0.001$; $n=20$), mounting ($r=0.99$; $p<0.001$; $n=20$) and fights ($r=0.87$; $p<0.001$; $n=20$). Play and fight behaviours showed low repeatability at short term (respectively $r=0.37$; $p=0.02$; $n=39$ and $r=0.35$; $p=0.03$; $n=39$). Play behaviours showed low consistency at medium term ($r=0.26$; $p=0.11$; $n=39$) while fights were not consistent at medium term ($r=0.04$; $p=0.82$; $n=39$). Social licking and mounting showed medium repeatability at short term (respectively $r=0.50$; $p=0.001$; $n=39$ and $r=0.52$; $p<0.001$; $n=39$) and at medium term (both $r=0.53$; $p<0.001$; $n=39$). When the level of play and social licking increased between week 9 and week 13 (respectively $p=0.02$ and $p=0.04$), the level

of mounting and fighting decreased (respectively p=0.02 and p=0.03). Observations were linked with the activity level of the group (except for mounting were only a tendency was found). The more calves were active (standing) the more they were performing play, fights and social licking.

Social licking is an acceptable candidate for welfare assessment. The observation is feasible, its validity is acceptable and the measure shows good correlations between observers and acceptable correlations between days. The behaviour of the calves is evolving between weeks 9 and 13. Therefore the observation should be performed at a fixed age to be valid. Play behaviours, mounting and head-to-head fights have good repeatability between observers but low repeatability at short and medium term. They don't seem to be good candidates.

20.2 INTRODUCTION

Play behaviours and social interactions are natural behaviours in calves. Social licking between calves is, like similar to other animal species, associated with pleasurable or calming sensations for the actor and/or the receiver of this behaviour (see, for example, the study by Feh et al., 1993, showing that grooming lowered heart rate in horses), and with an increased performance such as a higher weight gain (Sato et al., 1984). Plays (locomotor movements other than walking or mounting) can be composed of a large range of behaviours like running, buck kicking and jumping. They are natural behaviours implicated in the development of young animals.

In contrast to a number of 'negative' indicators of veal calf welfare such as those related to fear of humans or novel objects, positive indicators of welfare have not been validated in terms of predictive or concurrent validity. Intuitively, these behaviours could be considered as positive indicators of veal calf welfare. However, for each of these behavioural categories, high levels of behaviour may be associated with high levels of welfare, but may also result from a certain deficiency in the environment and, therefore, reflect a form of redirected behaviour or a behavioural rebound following deprivation. For example, veal calves may perform high levels of allo-grooming (labelled 'excessive grooming' by some authors) because of inadequate feeding or boredom (e.g., Terosky et al., 1997; Fraser and Broom, 1990). Likewise, calves kept in isolation or with a reduced space allowance during rearing, exhibit an increased locomotor activity, including running and buck kicking, when taken out of their home environment into a test arena (Dellmeier et al., 1985; Jensen, 1999). In both of these latter examples, high levels of respective behaviours seem to represent indicators of reduced rather than increased welfare. Other social interactions like fights or mounting can also have different impact on the welfare of calves depending on their importance. For example, excessive mounting can cause injuries.

To be complete, welfare assessment should include these natural and 'positive' behaviours like social behaviours and play and should also assess the level of other natural behaviour like fights or mounting which can reveal welfare issues. The development of welfare assessment scheme creates then a need in developing methods to assess positive and natural behaviours on farm. This method should permit to evaluate with simple tests or observations the level of positive and natural behaviours and be efficient (i.e. determining as much as possible the real level of expression of these behaviours). To create a valid method four conditions must be completed.

- Is the measurement feasible on farm conditions? (i.e. Is the duration compatible with on-farm use?) It is possible to perform it in any farm, type of building?)
- Is the measure valid? (i.e. Does it tell us about the actual level of welfare of the group and is the level of behaviour influenced by environmental factors?)
- Is there an observer effect?
- Is the measure consistent in time?

Feasibility of the recording of social licking requires special attention, since previous research suggests that the incidence of this behaviour is generally low. For example, a study on social licking in heifers and steers (Sato, 1984), reported that in groups of 24 animals on average 15 social licking interactions took place, lasting around 40 seconds, over a total observation time of 18 hours divided across different days. The average time an animal spent licking was about 25 seconds per hour. Another study performed by Raussi et al. (2003) in calves, determined 'contacts with neighbours' for calves housed individually (but having the possibility to interact with other calves through the walls of the pen) or housed in pairs. On average, in only 2% of the scans (every 5 minutes during 12 hours) this particular behaviour (including close contact, licking etc.) was observed. This very low incidence may compromise feasibility of this parameter under commercial conditions. Play and other behaviour also have low incidences which makes their level of expression measurement difficult.

This experiment aimed at building a method to asses the level of positive and natural behaviours in group-housed calves and to determine its validity in regard with the four criteria cited above. It did not address the problem of validity of putative positive indicators of veal calf welfare nor will address matters concerning the interpretation of measurements in terms of welfare.

20.3 METHODS

As presented above, four categories of measures were distinguished related to positive and natural behaviours:

- Play behaviours: jumping, kicking, running…;
- Social licking;
- Mounting;
- Head-to-head fights.

In order to perform correctly the observations under commercial conditions, observers were trained based on earlier observations recorded on video. Initial observations over longer periods of time on a few farms were analysed in order to try to identify more limited time frames with high probabilities for the occurrence of social licking and play behaviour.

On farm observations were performed on 20 farms in France between October 2005 and March 2006. Calves were housed in large groups (between 20 and 64 calves per pen) and fed with an automatic milk dispenser (AMD). Farms were characterised by two types of milking systems: in instantaneous systems the milk was available all day long, in session feeding systems the milk was only available at certain times (most of the time 2 sessions of 6 hours per day).

Two randomly chosen pens were observed per farm (for a total of 39 pens observed). Observations were performed simultaneously by two trained observers. Observations were performed on 4 days: 2 observations at a two day interval at week 9 (9 weeks after the arrival of the calves in the farm) and 2 observations at a two day interval at week 13. Three observations were performed per day (morning, noon and afternoon). A total of 12 observations were performed per pen.

Observations of play behaviours, mountings and fights lasted 30 minutes separated in 15 periods of 2 minutes. Within each 2 minutes period, the number of calves performing the behaviours (play or mount or fight) was recorded. The decision was taken to record the number of calves performing each behaviour instead of the occurrence of each behaviours to avoid the possibility of high levels of behaviours due to one calf playing/mounting/fighting frequently during the observation. The data produced were mean proportions of calves playing/mounting/fighting for the 30 minutes sessions (mean for the 15 periods).

For social licking, the sampling method used was scan sampling (with a two minute interval). Each observation lasted 30 minutes (16 scans per observation). As not all animals were visible for each scan, the number of calves observed per scan was recorded for each scan. Data provided were mean proportions of observed calves performing social licking.

20.3.1 Data Analysis

Inter-observer repeatability was assessed through spearman rank correlations.

Between-day consistency was assessed at short term (day1–day2 and day3–day4) and at medium term (week9–week13) using spearman rank correlations.

Analyses of variance were used to assess the effects of environmental parameters (size of the group, feeding system…) on the observation and to analyse farm and pen effects. The model included:

- Fixed variables : day of observation (or week of observation), period of observation, moment of the feeding session, number of calves in the pen, breed of the calves;
- Random variable : farm;
- Interactions: farm and week, farm and period, week and period, week and number of calves, period and number of calves.

20.4 RESULTS

The four behaviours were not frequently observed. The mean percentages per observation of calves performing the behaviours were 0.99% ± 0.05 for play, 1.77% ± 0.05 for social licking, 0.45% ± 0.03 for mounting and 1.41% ± 0.07 for fights. Despite this rarity of occurrence, all four measures were affected by an important farm effect ($p<0.001$) showing that these measures permit to distinguish between farms. Observers agreed on the level of plays ($r=0.95$; $p<0.001$; $n=20$), social licking ($r=0.90$; $p<0.001$; $n=20$), mounting ($r=0.99$; $p<0.001$; $n=20$) and fights ($r=0.87$; $p<0.001$; $n=20$). Play and fight behaviours showed low repeatability at short term (respectively $r=0.37$; $p=0.02$ and $r=0.35$; $p=0.03$). Play behaviours showed low consistency at medium term ($r=0.26$; $p=0.11$) while fights were not consistent at medium term ($r=0.04$; $p=0.82$). Social licking and mounting showed medium repeatability at short term (respectively $r=0.50$; $p=0.001$ and $r=0.52$; $p<0.001$) and at medium term (both $r=0.53$; $p<0.001$). When the level of play and social licking increased between week 9 and week 13 (respectively $p=0.02$ and $p=0.04$), the level of mounting and fighting decreased (respectively $p=0.02$ and $p=0.03$).

Observations were linked with the activity level of the group (except for mounting where only a tendency was found). The more calves were active (standing) the more they were performing play, fights and social licking. All measures were influenced by the period of observation with calves performing more of these behaviours in the morning than at noon or in the afternoon. Play and fight behaviours were influenced by the moment of feeding session with calves performing less behaviours at the beginning of the feeding session. We are not sure if these two effects are direct or indirect (due to different levels of activity). Additional work will consider the effect of the activity level on positive behaviours (for instance, expressing the data as proportions of active calves). A pen effect was found for social licking and mounting for respectively 21% and 16% of the farms. For the observation of social licking, the activity level and the size of the group influenced the mean percentage of calves that could be seen by the observers ($r=-0.19$; $p<0.001$; $n=455$ and $r=-0.09$; $p=0.03$; $n=455$).

20.5 CONCLUSIONS

Social licking is an acceptable candidate for welfare assessment. The observation is feasible, its validity is acceptable and the measure shows good correlations between observers and acceptable correlations between days. The behaviour of the calves is evolving between week9 and week13. Therefore the observation should be performed at a fixed age to be valid. Play behaviours, mounting and head-to-head fights have good repeatability between observers but low repeatability at short and medium term. They are rare behaviours and additional works may have to consider them as binary data (absence/presence of the behaviour). They don't seem to be good candidates.

21

ASSESSMENT OF GENERAL FEARFULNESS IN VEAL CALVES HOUSED IN LARGE GROUPS

H. Leruste, J. Lensink and C.G. van Reenen

21.1 SUMMARY

The concept of ‘general fearfulness’ is quite controversial (Ramos and Mormède, 1998; Wilson, 1998). Whether or not the fear of novel objects reflects ‘general fearfulness’ remains to be determined but we can assume that the behavioural response of veal calves towards a novel object represents a valid measure of fear. The reaction of calves towards on novel object can be easily assessed on experimental conditions. Under commercial conditions only few studies have been performed so far on feasibility and reliability of these measures. This study aims at determining simple and valid tests which would be good candidates to integrate an animal-based-on-farm assessment tool.

Observations were performed on 20 farms in France. Calves were housed in large groups and fed with an automatic milk dispenser. Observations were performed on 4 days: 2 observations at a two day interval at week 9 (9 weeks after the arrival of the calves in the farm) and 2 observations at a two day interval at week 13. Observations were performed simultaneously by two trained observers. Two observations were performed. The voluntary approach of a novel object in the middle of the pen permitted to assess the latency for the group to touch and surround the object and the percentage of calves orientated towards the object. The voluntary approach of a novel object at the front fence permitted to assess the mean number of calves around the object during 3 minutes and to build a speed-of-approach score for the group.

A farm effect was found for all measures showing that they permit to distinguish between farms. Observers agreed for all the measures evaluated ($r>0.75$) and all measures had satisfying repeatability at short term ($r>0.56$). The percentage of calves orientated toward the object in the pen, the mean number of calves close to the object at the fence and the speed-of-approach score showed good repeatability at medium term (respectively $r=0.61$; $p<0.001$; $n=54$ / $r=0.60$; $p<0.001$; $n=67$ / $r=0.67$; $p<0.001$; $n=60$). The latency to touch and to surround the object in the pen showed low repeatability at medium term (respectively

r=0.34; p=0.02; n=48 and r=0.47; p=0.001; n=44). We observed a habituation of the calves to object in the pen showed by the decrease of the latency to surround the object between day1 and day4 (p=0.009). Calves approached faster the object at the fence and were more numerous close to it at week 13 than at week 9 (p<0.001). Therefore observations should be performed at a fixed age.

This experiment permitted to develop two observations (providing three variables) for on-farm assessment of general fearfulness: the latency to touch and to surround a novel object in the pen and the mean number of calves around a novel object at the fence. Despite this, the use of these measures is not recommended because of their low feasibility. They require an adaptated equipment and can be affected by the presence of the observers.

21.2 INTRODUCTION

The concept of 'general fearfulness' is quite controversial (e.g. Ramos and Mormède, 1998; Wilson, 1998), and a number of studies in veal calves and cattle suggest that, for example, fear of humans may be dissociated from fear of novel objects (Hemsworth et al., 1996; Lensink et al., 2000; Van Reenen et al., 2004). Moreover, there is strong disagreement on the interpretation of various behavioural measures observed during intuitively fearful situations such as an 'open field test' or a 'novel object test'. For example, based on work with dairy cattle, it was suggested that individual differences in latencies to approach humans or objects as well as differences in locomotion and vocalisation during novel object, human approach, and open field tests were all associated with differences in fearfulness (Boissy and Bouissou, 1995; De Passillé et al., 1995; Hemsworth et al., 1996; Grignard et al., 2000). However, multivariate analyses of behavioural measures simultaneously recorded during behavioural tests in cattle, but also in other species like pigs or rats, usually reveal a multifactorial picture in that different behaviours load on different factors (e.g. Andersen et al., 2000; Forkman et al., 1995; Grignard et al., 2001; Steimer et al., 1997; Van Reenen et al., 2004). This strongly suggests that the behavioural responsiveness of animals, including veal calves, to challenge is mediated by a number of factors or motivations rather than a unidimensional construct like general fearfulness.

A pharmacological validation study demonstrated that the treatment of calves with an anxiolytic drug (a benzodiazepine labelled brotizolam©) did not affect locomotion or vocalisation during an open field test, but specifically increased the time spent interacting with the stimulus (a plastic container attached to a tambourine) in a novel object test (Van Reenen et al., 2009). These findings support the previously mentioned multidimensionality of responsiveness of calves to challenge, and they also provide evidence for predictive validity of this latter behavioural parameter as a measure of fear in calves. The interaction of calves with a novel object also seems to comply with the condition of concurrent validity, since cortisol responses to open field and novel object tests (i.e. putative physiological indices of stress and fear) were positively associated with the latency to approach a novel object, and negatively related to the time spent in contact with the novel object (Van Reenen et al., 2005).

Whether or not this type of fearfulness reflects ‘general fearfulness’ remains to be determined but we can assume that the behavioural response of veal calves towards a novel object represents a valid measure of fear. Relatively simple and feasible behavioural measures in veal calves in the area of the fear of a novel object have been sufficiently validated. It can be easily assessed on experimental conditions. Only few studies have been performed so far on feasibility and reliability of these measures under truly commercial conditions. The development of welfare monitoring schemes (EU project) creates a need for a valid method to assess general fearfulness on farm.

This study aims at determining simple and valid tests which would be good candidates to integrate an animal-based-on-farm assessment tool. They should reflect the actual level of fear/curiosity of a novel object of the group of calves, be repeatable at short and medium term (test-retest) and shouldn’t be affected by an observer effect.

21.3 METHODS

As presented above, two categories of measures were performed:

- voluntary approach of an object in the middle of the pen;
- voluntary approach of an object at the front fence.

On farm observations were performed on 20 farms in France between October 2005 and March 2006. Calves were housed in large groups (between 20 and 64 calves per pen) and fed with an automatic milk dispenser (AMD). Observations were performed on 4 days: 2 observations at a two day interval at week 9 (9 weeks after the arrival of the calves in the farm) and 2 observations at a two day interval at week 13. Between 1 and 6 pens were observed per farm (for a total of 67 pens observed). Observations were performed simultaneously by two trained observers.

The two observers calmly entered the building. One observer entered the pen with the object and placed it in the middle of the pen in a standardised manner. The object used was unfamiliar to the calves. It was nearly 1,30m high with a heavy base and a coloured top. The two observers noted down the latency for the first calf to touch the object and the latency for the object to be surrounded by calves (no space left for other calves to approach) with a maximum latency of 180s. At the end of the test the observers evaluated the percentage of calves that were orientated towards the object. The second observation was performed just after the first one. An unfamiliar and coloured object was placed at the front fence of the pen in a standardised manner. The two observers counted every 20 seconds (10 scans) the number of calves in the area close to the object (1.5m).

21.3.1 Data Analysis

Inter-observer repeatability was assessed through spearman rank correlations. Between-day consistency was assessed at short term (day1–day2 and day3–day4) and at medium term (week9–week13) using spearman rank correlations.

Analyses of variance were used to assess the effects of environmental parameters (size of the group, breed, order effect) on the observation and to analyse farm and pen effects. The model included:

- fixed variables: day of observation (or week of observation), number of calves in the pen, breed of the calves and order of performance of the test;
- random variable: farm;
- interactions: farm and day (or week), number of calves and day (or week) and order and day (or week).

The evolution of the variables in time was assessed thanks to a repeated measurement analysis.

21.4 RESULTS

A farm effect was found for all measures showing that they permit to distinguish between farms. Observers agreed for all the measures evaluated: latency to touch the object in the pen ($r=0.80$; $p<0.001$; $n=20$), latency to surround the object in the pen ($r=0.75$; $p<0.001$; $n=20$), % of calves orientated toward the object ($r=0.90$; $p<0.001$; $n=20$), mean number of calves close to the object at the fence ($r=0.81$; $p<0.001$; $n=20$) and speed-of-approach score ($r=0.84$; $p<0.001$; $n=20$). All measures had satisfying repeatability at short term (day1-day2): latency to touch the object in the pen ($r=0.67$; $p<0.001$; $n=33$), latency to surround the object in the pen ($r=0.56$; $p=0.002$; $n=29$), % of calves orientated toward the object ($r=0.62$; $p<0.001$; $n=48$), mean number of calves close to the object at the fence ($r=0.64$; $p<0.001$; $n=61$) and speed-of-approach score ($r=0.61$; $p<0.001$; $n=44$). The percentage of calves orientated toward the object in the pen, the mean number of calves close to the object at the fence and the speed-of-approach score showed good repeatability at medium term (respectively $r=0.61$; $p<0.001$; $n=54$ / $r=0.60$; $p<0.001$; $n=67$ / $r=0.67$; $p<0.001$; $n=60$). The latency to touch and to surround the object in the pen showed low medium term repeatability (respectively $r=0.34$; $p=0.02$; $n=48$ and $r=0.47$; $p=0.001$; $n=44$).

We observed a habituation of the calves to object in the pen showed by the decrease of latency to surround the object between day1 and day4 (p=0.009). Calves approached faster the object at the fence and were more numerous close to it at week 13 than at week 9 (p<0.001). Therefore, observations should be performed at a fixed age.

A pen effect was found for all measures. It was more important for the test of the object at the fence (47% of the farms were affected by this effect). Therefore the tests should be performed on at least two pens. There is an effect of the order of performance of the test in the different pens on the latency to touch the object (p=0.005), the percentage of calves orientated (p=0.009) and the speed-of-approach score (p=0.011). Therefore, the chosen tests should be adjacent to diminish this effect. The breed of the calves didn't influence their behaviour. The size of the group influenced the percentage of calves orientated towards the object (p=0.08) and the speed-of-approach score (p=0.047).

All the measures are linked. Groups which show short latency to touch the object in the middle of the pen had also short latency to surround the object and were more orientated toward the object. They also were more numerous to surround the object at the fence and approached faster.

21.5 CONCLUSIONS

The latencies to touch the object and to surround it show low between-day repeatability due to an evolution of the behaviour of the calves. Therefore it can't be truly assessed. The mean number of calves close to the object at the fence and the speed-of-approach score show good between-day and between-observers repeatability.

This experiment permitted to develop three valid observations for on-farm assessment of general fearfulness: latency to touch and to surround a novel object in the pen and mean number of calves around a novel object at the fence. Despite this, the use of these measures is not recommended because of their low feasibility.

22

ASSESSMENT OF GENERAL FEARFULNESS IN DAIRY COWS AND FATTENING BULLS

S. Waiblinger and I. Windschnurer

22.1 SUMMARY

Fearfulness is an individual's propensity to be easily frightened or alarmed in a variety of potentially aversive situations (Boissy, 1998; Erhard and Schouten, 2001). Thus, if fearfulness is high, the animals will experience the negative emotional state of fear more often. In cattle husbandry most potentially fear-eliciting events are associated with humans and with novelty (e.g., new feeding equipment or other barn structures when regrouped). Hence, measuring states of fear associated with those events seems to be most relevant for welfare. The level of fear of humans is included in the assessment of the animals' relationships to humans.

After some trials, where several possibilities were evaluated for on-farm feasibility, two tests were further investigated: The 'novel object test' (NOT) is based on approach behaviour of a group of animals to a novel object (a ball hanging at the feed barrier). The 'vigilance test' (VT) is based on the frequency of vigilant behaviour (scanning the environment) during feeding, also in presence of a novel object (the camera filming the reactions of the animals).

The NOT was evaluated with regard to inter-observer reliability, test-retest repeatability, and convergent validity on 16 dairy farms (where two tests at two different places were performed on every farm) and on 10 bull fattening farms (where 2 to 15 pens were tested per farm). The vigilance test (VT) was evaluated with respect to test-retest repeatability and convergent validity in dairy cows. Two groups of cows were filmed during feeding and their behaviour categorised as vigilant, vigilant and looking into the camera or not vigilant.

Inter-observer reliability was high for the novel object test in dairy cow and fattening bull farms ($r=0.95$ to 0.98, $p<0.001$, $N=16$ / 49, farm / pen level). The test was not repeatable between two test sessions performed in sequence ($r \leq 0.2$). Regarding convergent validity, measures of the first test correlated especially with measures indicating a higher level of

fear of humans in dairy cows (farm level: r= –0.49 to –0.62, p<0.05, N=16). In fattening bulls, correlations with the average avoidance distance at the feeding place were found (pen level: r= –0.31; p<0.05; N=49 or r= –0.24; p<0.1; N=49). As with the novel object test, no repeatability could be found between the two tests (two 4 min records of focal groups) (herd level: %vigilant animals, 1st and 2nd test: r=0.14, p=0.642, N=15). The test measure '%vigilant animals in the first test' correlated significantly with the proportion of animals very fearful of humans at the feeding place (herd level: r=0.55, p<0.05, N=15). The two fear tests showed few correlations with each other.

In sum, there are some indications for the validity of measuring vigilance, but at the moment the measure is too premature to include it into a welfare assessment protocol. The novel object test seems to have some validity with respect to measuring general fearfulness in the herd.

22.2 INTRODUCTION

Fear is a negative emotional state. It has welfare relevance (1) if the animal experiences states of fear or anxiety for a prolonged time and/or frequently and (2) due to the potential negative consequences of fear reactions such as injuries. Fearfulness, being a personality trait, is an individual's propensity to be easily frightened or alarmed in a variety of potentially aversive situations (Boissy, 1995; Erhard and Schouten, 2001). It can have a considerable influence on the animals' physiological or behavioural reactions. From this definition fearfulness cannot be measured with a single test, as showing fear in one certain situation is considered just as a state (Erhard and Schouten, 2001). However, it is not feasible to include a series of time-consuming tests into an on-farm welfare assessment tool. In cattle husbandry fear-eliciting events are most likely associated with humans (with or without handling the animals) and with novelty (of the environment or situation, e.g., new feeding equipment when regrouped). Thus, measuring states of fear associated with those events seem to be most relevant. The level of fear of humans is assessed by the assessment of the animals' relationships to humans (Chapters 15–17). We aimed at measuring the level of fear in another specific test, also assuming that it is nevertheless easier to create a state of fear in animals that are generally more fearful.

After evaluating several tests with respect to feasibility we finally chose a novel object test for assessing general fearfulness in dairy cows as well as in fattening bulls, since novelty is known to be a frightening stimulus (e.g., Boissy 1995). Novel object tests are one of the most commonly used tests to assess fear in cattle (Forkman et al., 2007), although according to Forkman et al. more work is necessary to investigate its validity. In addition, in dairy cows we evaluated possibilities to base a measure of fear on vigilance during feeding because Welp et al. (2004) suggested that measuring vigilance in dairy cows could be used as an indicator of the degree of fearfulness. Animals tested individually in an experimental set-up spent more time being vigilant in the presence of a dog than a human and in presence of an aversive human compared to unfamiliar or gentle people.

22.3 METHODS

Before the actual on-farm study with 16 dairy cow and 10 bull fattening farms several tests were assessed for their feasibility. At first, two fear tests were performed at an individual level in one commercial dairy farm with loose housing system. In one test single cows had to pass between two novel objects that were placed in a corridor leading from the milking parlour back to the stable. The other test was a surprise test (air blow test, inspired by Schrader, 2002) carried out with individual animals restrained in the feed barrier. 45 and 46 cows were subjected to the two tests and their reactions were filmed. However, both tests turned out to be not feasible for on-farm welfare assessment due to the difficulties to standardise the procedure, the time needed, the lack of applicability in different housing systems (e.g., without possibilities to restrain the animals in the feed barrier) and the necessity of close human involvement for performing the test. Therefore, the two tests were not further evaluated.

The test supposedly most feasible and most easily performed with low human involvement and possible for dairy cows as well as fattening bulls was the presentation of a novel object at the barn barrier. Therefore, it was decided to evaluate a novel object test (NOT) with respect to test-retest repeatability, inter-observer reliability, and convergent validity (for definitions see Chapter 16). However, it could only be evaluated at farm and at pen level but not at an individual level. This was because an individual presentation of a novel object was not possible for many animals since neighbouring animals would have inevitably seen the object too and it would not have been novel anymore. We also considered other possibilities for inclusion of an element of 'suddenness' into the test but failed to find a practicable solution (due to constraints of on-farm assessment). In addition to the novel object test, we made a first investigation into the repeatability and validity of vigilance measures in dairy cows because recent papers suggest it to be indicative of fear. In order to investigate convergent validity correlation between general fearfulness tests and tests assessing the animals' relationship to humans were evaluated

22.3.1 Farms, Animals, and Housing

Dairy Cows

For testing the reliability and convergent validity of the two tests chosen for dairy cows, 16 commercial dairy farms with 19 to 78 lactating cows (36 ± 15 per herd, in total 580 cows) were visited. Depending on the tests that were performed, different numbers of animals out of the herd were tested. On one farm, cows were housed in a deep litter system and on the other farms in cubicle loose housing. The running area was made of slatted concrete floors on 11 farms and of solid concrete on 5 farms. The farms kept predominantly Simmental cows, with a few Holstein Friesian cows on some of the farms. One farm kept Holstein Friesian cows only.

Fattening Bulls

10 farms with altogether 123 pens (housing 3 to 12 bulls, on average 6 animals per pen) were visited. Farms had between 8 and 18 pens and 40 to 123 bulls (73 ±24). On one farm bulls were housed in deep litter, on 9 farms bulls were housed on fully slatted concrete floors (except for two pens with straw bedding). The weight classes ranged from estimated 200kg up to finishing bulls with approximately 700kg. The animals were mostly Simmental bulls, the breed traditionally found on commercial Austrian bull fattening farms.

22.3.2 Description of Test Procedures and Measures Calculated

For dairy cows two measures were evaluated, a novel object test and a measure of vigilance. For fattening bulls, the novel object test was tested.

Novel Object Test (NOT)

The novel object (a coloured plastic ball of ~25 cm in diameter for the dairy cows and ~30 cm for the fattening bulls (different equipment for hygienic reasons) was knotted into a plastic cord to be easy to disinfect. For each test the ball was attached to the feed barrier within good view for many animals of the herd in dairy cows or for the whole pen in fattening bulls (pens having smaller group sizes). Before the attachment of the ball a position was selected for the ball and then the distances of 5 m to the right and the left of this position were marked outside the feed barrier (with tapes, straw, hay, and feed). Then two observers took position in a distance of at least 5 m from the novel object. One of them alternatively attached the ball to the feed barrier and then went back to the observer position. Both observers noted the animals' reactions for the duration of 5 minutes.

The following data were collected independently by each observer:

- Number of animals in the whole group (pen in fattening bulls, herd in dairy cows) at test start
- Number of animals standing within 5m to the novel object at test start for dairy cows, standing in the pen for fattening bulls
- Number of animals approaching until contact

From the collected data percentages of animals approaching until contact were calculated relative to the number of animals standing within 5 m at test start (for dairy cows) or relative to the number of animals standing in the pen (for fattening bulls) at test start or relative to the number of animals in the whole group. The observers tried to note the identity of the animals (ear tags) in order to avoid recording the contact made by individual animals more than once.

Vigilance Test (VT)

A group of feeding animals that was not restraint in the feed barrier was filmed angular ahead using a video camera positioned on a tripod for 5 to 6 minutes. After starting the video tape, the humans left the barn, so that the animals were undisturbed by humans. The number of animals visible on the video in such a focal group ranged from 0 (in few scans when animals had left places) to 9 animals.

The behaviour of the animals was recorded from video by scan sampling, with scans every 10 seconds for 4 min, starting 50 sec after the person had started the video. The first 50 sec were disregarded to be sure that the person was out of the view of the animals. Animal behaviour was categorised as follows:

- vigilant without looking to the camera (vigilant): head held up (line of the neck parallel to ground or in an angle >90°) & ears erect & immobile; the definition is based on Boissy and Dumont (2002) and Mounier et al. (2006)
- vigilant or attentive looking into the camera (vigilant to camera): direct look into direction of camera, head held up or lowered
- not vigilant: lowered head (line of neck in an angle <90° to ground), not looking to the camera, occupied with feeding or social interactions; also animals entering or leaving the feed barrier at the moment of the scan were put into this category

Only animals with their heads reaching completely through the feed barrier were counted. To be able to differentiate vigilant animals from animals lifting the head to leave or enter the feed barrier, the seconds before and after the scan were watched.

The average percentage of animals in each category was calculated.

22.3.3 Design of the Study

The measures for general fearfulness were tested for reliability together with measures for assessing the animal-human relationship during the same farm visit (see Chapter 15).

Novel Object Test (NOT)

The novel object test was always performed as last test after the tests for assessing animal-human relationship (AHR) had been performed. In dairy cows 6 AHR tests were carried out: Avoidance distance at the feeding place, Approach test to a stationary person, Lypass, Lytouch, Walking trough and touch, and the Avoidance distance test in the barn. In fattening bulls only the Avoidance distance in the feeding place was carried out. For exact description of these HAR tests see Chapters 16 and 17.

In dairy cows, the novel object test was repeated two times with a short break of 5 minutes in between, when the novel object was placed in a different position which had to fulfil the same criteria as the first place (visibility for animals etc.).

In fattening bulls the test was performed once with 2 to 15 (on average 6) pens per farm. In total 49 pens were fully tested. For another 10 pens the test had to be interrupted to avoid the destruction of the novel object by the bulls, or because the holding cracked or the farmer entered the barn and disturbed the animals. Repetition within farms was balanced for age. However, from the fourth farm on the test was performed only with the lighter weight classes because the heavy bulls (1) often played very wildly so that the holding cracked or they even nearly destroyed the ball and (2) hardly showed signs that could be interpreted as fear.

Vigilance Test (VT)

Vigilance was assessed two times per dairy herd with presumably different focal groups of cows by changing the position of the camera between the two tests without a pause in between. To guarantee a sufficient number of animals still feeding VT was performed after releasing the cows from the feed barrier (if restrained) after they had been subjected to the avoidance distance at the feeding place (AHR test).

22.3.4 Data Analysis and Statistics

Convergent validity and reliability (inter-observer reliability and test-retest repeatability) were analysed using Pearson or Spearman rank correlations, depending on the distribution of the data.

22.4 RESULTS

22.4.1 Validity – Inter-test Correlations

Novel Object Test

A large variation was found in behaviour on the 16 dairy cow farms. The proportion of cows approaching the novel object until contact ranged from 3.3 to 15.8 % (median: 7.7%; average ± stand.dev.: 9±4%) in the first test session and from 2.6 to 23.9 % (median: 9.9%; average ± stand.dev.: 11±6%) in the second.

The proportion of cows standing within 5 m distance at test start and approaching until contact ranged from 23 to 100 % (median: 55%; average ± stand.dev.: 56±21%) in the first and from 33 to 100 % (median: 59%; average ± stand.dev.: 65±23%) in the second test session.

With regard to inter-test relationships with AHR tests, the first novel object test showed significant correlations with the % of animals having an avoidance distance >1.5m in the barn or >50cm and >1m at the feeding place (Table 22.1), i.e., with the number of very

Table 22.1 Pearson or Spearman correlations of the Novel object test parameters (averages of both observers for first and second test session) with measures assessing the animal-human relationship (always first test session) (N=16, except where indicated N=15).

	% to contact from 5m at test start – 1st test	% to contact from 5m at test start – 2nd test	% contact from animals per barn – 1st test	% contact from animals per barn – 2nd test
Approach to person				
% app to contact out of 10m at test start	.21	.04	–.18	**–.59***
% app to contact out of standing animals in the barn (N=15)	.00	–.28	.38	***.45***
Avoidance distance in the barn (AD)				
Average AD	–.20	–.42	–.32	–.35
ADtouch	.28	***.44***	.20	.08
AD>50cm	–.22	***–.53****	–.20	–.32
AD>1.5m	–.21	–.14	**–.53***	–.38
AD>2m	–.15	–.10	–.39	–.16
Avoidance distance at feeding place (ADF)				
Average ADF	–.28	**–.52***	–.39	–.41
ADF% touched	.07	**.72****	.02	.38
ADF% > 20 cm	–.22	**–.61***	–.30	***–.47***
ADF% > 30 cm	–.31	**–.55***	–.39	–.41
ADF% > 50 cm	–.36	–.37	**–.51***	–.36
ADF% > 1 m (Spearman)	–.20	–.04	**–.62***	–.18

Notes: all p<0.1 are bold; in italics: p<0.1; * p< 0.05; ** p< 0.01; *** p<0.001.

fearful animals in the avoidance distance tests. In contrast, the second novel object test correlated significantly with the other measures of avoidance distance (lower distances and average) and with the number of animals approaching the test person. One correlation was especially high: the number of animals approaching the novel object until contact (in relation to the number of animals standing within 5 m distance) was higher the more animals could be touched at the feeding place (Table 22.1).

In fattening bulls, a farm effect could be detected by ANOVA when only including the younger animals, i.e., pens with weight classes up to 300 kg ($p = 0.02$, $N= 17$ pens from 6 farms), but not when including all weight classes. Correlations with the measure of the animal-human relationship were found. The percentage of bulls approaching the novel object until contact relative to the number of animals standing at test start correlated significantly with the average avoidance distance (average ADF) of the second test session ($r=-0.31$; $p<0.05$; $N=49$) and only tended to correlate with the average ADF of the first test session ($r=-0.24$; $p<0.01$; $N=49$).

Vigilance

The vigilance measures varied widely between farms. The number of vigilant animals (%vigilant) ranged from 0 to 66 % (median: 14%; average ± stand.dev.: 21±20%) in the first and from 4 to 48 % (median: 14%; average ± stand.dev.: 16±11%) in the second test session.

In the first test session, the % of vigilant animals correlated significantly with the % of animals having an avoidance distance >1m at the feeding place (Table 22.2), i.e., with the number of very fearful animals in this avoidance distance tests. No significant correlation was found for the second test session (Table 22.2).

No consistent pattern of correlations and only two significant correlation coefficients were found between measures of vigilance and the novel object test (Table 22.3).

22.4.2 Reliability

Novel Object Test (NOT)

A high inter-observer reliability was found for all measures both in dairy cows (ranging from $r=0.95$ to 1.00) and fattening bulls (average latency: $r=0.96$; %contact of all bulls in pen: $r=0.98$; % contact of standing bulls: $r=0.98$, all $p<0.001$, $N=49$; Figure 22.1).

No to extremely low repeatability between test sessions (test-retest repeatability) was found in dairy cows: correlation coefficients of first and second test $r \leq 0.2$.

TABLE 22.2 Pearson or Spearman correlations of the percentage of animals being vigilant or attentive looking at the camera, being vigilant without looking at the camera or being not vigilant (each for the first and the second test session) with measures assessing the animal-human relationship (always first test session).

	% vigilant to camera, 1st test N = 15	% vigilant, 1st test N = 15	% not vigilant, 1st test N = 15	% vigilant to camera 2nd test N = 14	% vigilant, 2nd test N = 14	% not vigilant, 2nd test N = 14
Approach to person						
% app to contact out of 10m at test start	–.28		.13			
% app to contact out of standing animals in the barn	–.42	–.14	.19	–.35	.37	–.19
Avoidance distance in the barn (AD)						
average AD	.25	.18	–.21		–.22	.20
ADtouch		–.15	.13			
AD>50 cm	.13				–.23	.22
AD>1.5m	***.45***	.30	–.35		–.26	.25
AD>2m	.43	.37	–.41		–.20	.14
Avoidance distance at feeding place (ADF)						
average ADF		.25	–.24		.16	–.12
ADF% touched	.26			–.19		.14
ADF% > 20 cm		.16	–.14	–.11	.16	
ADF% > 30 cm		.24	–.22	–.17	.25	–.16
ADF% > 50 cm		.29	–.29	–.14	.21	–.14
ADF% > 100 cm (Spearman)	.38	**.55***	***–.46***			

Notes: coefficients ≤ ±0.10 not depicted; all p<0.1 are bold; in italics: p<0.1; * p< 0.05; ** p< 0.01; *** p<0.001.

TABLE 22.3 Correlations of vigilance (VT) and novel object test (NOT).

VT \ NOT	N	% to contact from 5m at test start – 1st test	% to contact from 5m at test start – 2nd test	% contact from animals per barn – 1st test	% contact from animals per barn – 2nd test
% vigilant to camera, 1st test	15	–.21	.42	–.22	.51*
% vigilant, 1st test	15	–.09	.12	–.16	.34
% not vigilant, 1st test	15	.11	–.18	.18	–.40
% vigilant to camera, 2nd test	14	–.11	–.07	.34	.09
% vigilant, 2nd test	14	.53*	–.17	.14	.19
% not vigilant, 2nd test	14	–.44	.18	–.26	–.20

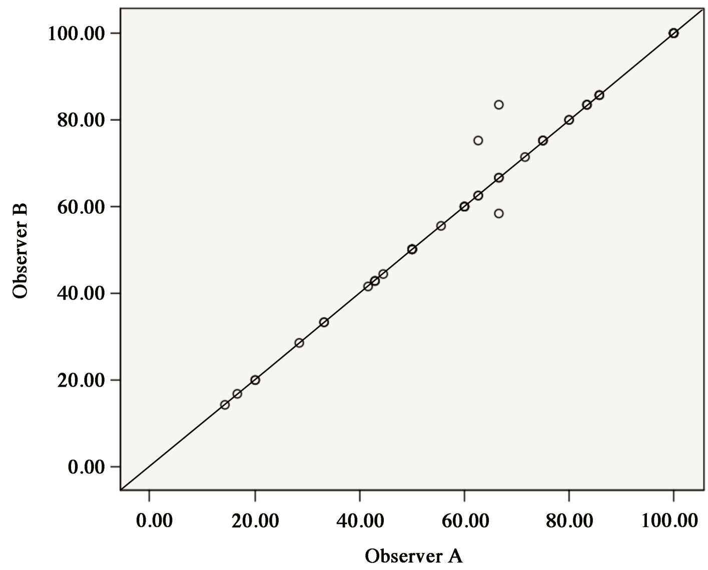

FIGURE 22.1 Scatter plot of the % of animals approaching to contact relative to the number of bulls per pen observed by person A and B simultaneously.
Notes: inter-observer reliability; Pearson correlation coefficient: r=0.98, p<0.001, N=49).

Vigilance

No repeatability could be found between the two test sessions (analysis of two 4 min records of two focal groups) (see Table 22.4). Depending on the measure, correlations with the average of both test sessions (e.g. average % vigilant) were high for the 1st test (e.g. % vigilant) or for the second test (e.g. % vigilant to camera, Table 22.4).

22.5 DISCUSSION

22.5.1 Novel Object Test

The measures showed very high inter-observer reliability both in fattening bulls and dairy cows. However, on the short term, the results were not repeatable, the first and the second test session in the dairy herds did not correlate. This may be an effect of habituation to and thus reduction in fear of the novel object (NOT) during the first test session (Forkman et al., 2007), both by exploring it or by watching others to explore it. Thus, the behaviour of the cows may reflect the level of fear or be influenced by fearfulness to a lower extent in the second test session compared to the first one. The correlations with the measures to assess the animal-human relationship (AHR) also point in this direction: In the dairy herds, the proportion of animals approaching the NOT relative to the number of animals standing

TABLE 22.4 Pearson correlation coefficients of the two tests of vigilance with each other as well as with the average of both (n = 14 farms).

	% vigilant, 1st test	% vigilant, 2nd test
Average % vigilant	.89***	.57*
% of vigilant cows, 1st test		.14
	% vigilant to camera, 1st test	**% vigilant to camera, 2nd test**
Average % vigilant to camera	.58*	.87***
% vigilant to camera, 1st test		.10
	% not vigilant cows, 1st test	**% not vigilant cows, 2nd test**
Average % not vigilant	.88***	.58*
% not vigilant cows, 1st test		.13

in the barn (% app NOT) in the first test showed significant correlations with measures reflecting a higher level of fear of humans in the herd (i.e., a higher proportion of very fearful animals in the avoidance distance tests). In contrast, for the second test sessions the proportion of animals standing within 5m approaching the NOT (% app NOT from 5m) correlated highly with the proportion of animals that could be touched at the feeding place. Furthermore, only the second test showed correlations with the approach behaviour to a stationary human, taking the measure % app NOT. These findings are not straightforward to interpret, but they suggest that the first test with % app NOT can give some indications of the level of fearfulness in the herd, i.e. if cows are easily frightened in response to humans and to novel objects. The second test and the measure % app NOT from 5m may be more influenced by other motivations and emotions, possibly the level of security felt when at the feeding place, which is the area of highest competition (see for the % cows that could be touched at the feeding place in Waiblinger et al., 2003), or by the level of curiosity and motivation to explore.

To evaluate the test-retest repeatability of a test using a novel object as stimulus is always difficult. The novelty may no longer exist after the first presentation. We assumed that by changing the position we were testing a second, different focal group that had not yet investigated the novel object. Even if the assumption was true for investigation in close proximity, we may have underestimated the ability to learn about potentially frightening stimuli from a distance and by social learning (Veissier et al., 1993) and thereby changing the level of fear.

In sum, our results support the notion that a novel object test gives some valid indications for the level of fearfulness in both loose housed dairy cows and fattening bulls. The test has nevertheless some problems that make it questionably for welfare assessment in our opinion. There are several confounding factors, i.e. curiosity, motivation to explore, feeding motivation, or social dominance interactions (dominant animals may prevent others to approach close), which may infer with validity. Observations of animals not approaching the object in order to get more information turned out to be too difficult.

Additionally, regarding the relevance of the test with regard to the welfare of cows it can be argued that its value added to the measures of avoidance distances may be not large enough to justify additional time in a welfare assessment protocol. As mentioned in the introduction, most possible fear-eliciting events are associated with humans, while confrontation with novel objects is rarer in cattle husbandry. Moreover, it is difficult to draw boundaries with respect to acceptable levels or level of welfare as there is no information available so far.

22.5.2 Vigilance

The test of vigilance was not repeatable. Also here a novel object, the camera, was an inherent part of the procedure and the animals may have habituated to it also from a distance leading to different reactions. Furthermore, we could not exclude testing at least partly the same animals twice.

Our hypothesis was that more fearful groups of animals would spend less time feeding but would scan their surroundings more often, i.e., spent more time being vigilant. We could test this only by correlations with other measures presumably measuring fear. The results suggest that assessing vigilance could have some validity to measure the level of fear. Comparable to the novel object test the proportion of vigilant animals in the first test correlated significantly with the proportion of animals very fearful of humans at the feeding place and thus may measure the level of fearfulness to some extent. However, few significant correlations with the novel object test were found.

Altogether, the picture is not consistent enough and thus can be seen only as indication that this measure may be promising also at herd level in the on-farm context. Taking results from the literature into consideration (Welp et al., 2004), we conclude that this measure may be useful. However, further investigations with respect to validity and repeatability are necessary. Another limitation for inclusion in a welfare assessment protocol is the feasibility. It is a disadvantage that video equipment is needed which (1) is difficult to clean and especially disinfect and (2) that additional analysis in an office is needed (scan sampling from videos).

22.6 CONCLUSIONS

For the vigilance test some indications were found for its validity, but at the moment the measure is too premature to include it into a welfare assessment protocol.

The novel object test seems to have some validity with respect to measuring general fearfulness in the herd, but still has some problems and – in our opinion – is less relevant for welfare compared to measures of the animal-human relationship (AHR), at least in dairy cows. Moreover, setting thresholds is difficult due to a lack of information. Thus, at the moment we do not recommend to include the novel object test in welfare assessment schemes but instead to give more time to a sound assessment of the AHR by avoidance distance tests.

23

QUALITATIVE BEHAVIOUR ASSESSMENT

F. Wemelsfelder, F. Millard, G. De Rosa and F. Napolitano

23.1 SUMMARY

Qualitative Behaviour Assessment (QBA) is a method that relies on the ability of human observers to integrate perceived details of behaviour, posture, and context into descriptions of an animal's style of behaving or 'body language', using descriptors such as 'relaxed', 'tense', 'frustrated' or 'content'. Such terms have an expressive, emotional connotation, and provide information that is directly relevant to animal welfare. Previous research with pigs, cattle, poultry, and horses has shown QBA to have high inter- and intra-observer reliability and to be coherent with traditional behaviour and welfare measures, both when animals were assessed individually and at group level. Previous QBA work however was based on a Free-Choice-Profiling methodology that is unsuitable for on-farm inspection work. The aim of this study therefore was to design, and test the inter-observer reliability of, a fixed rating scale for QBA of cattle expression. This work was carried out with beef cattle, dairy cattle and veal calves.

On the basis of previous QBA research with cattle and consultation with cattle experts, we designed a rating scale of 29 descriptors (32 for beef cattle). The rating scales were tested by three cohorts of four assessors, on 22 groups of veal calves and 22 groups of dairy cattle in Northern and Southern Italy, and on 21 groups of beef cattle in Southern Scotland. The inter-observer reliability of the QBA scores attributed to the different cattle groups was tested using Kendall Correlation Coefficient W. For beef cattle there was satisfactory reliability ($W \geq 0.70$) for 20 out of 32 descriptors, but for dairy cattle and calves this criterion was reached with only a few descriptors. Comparison of Principal Component Analyses (PCA) of assessor scores within the three cattle groups showed similar emergent patterns of cattle expression, in which the first principal component (PC1) distinguished between positive and negative mood, and the second (PC2) differentiated these moods in low and high levels of arousal. This pattern was reproduced when we removed descriptors with low loadings and low apparent welfare relevance from the assessor data sets, leaving 20 descriptors for each cattle group. For beef cattle, PC1 of the 'reduced' PCAs showed satisfactory inter-observer reliability (Kendall $W=0.73$; $p<0.001$, but for dairy cattle and veal calves these emergent patterns, though present, were quantitatively weak. A

subsequent video-based assessment of dairy cattle by 14 assessors using this rating scale found satisfactory reliability (W=0.73; p<0.001).

We propose that PC1 may provide an integrative measure of positive and negative cattle emotion, to be accorded to single farm units through a PCA of assessors' scores on a 20-term QBA rating scale. To calibrate the QBA measures of single farms, testing the QBA scale on a large sample of farm units is required to create a 'benchmark' data base. This will allow identification of cut-off points on PC1 for unacceptable levels of negative mood/welfare.

The application of QBA on farms is highly feasible and easy to learn, however assessors must be experienced in observing cattle, and be given additional training in recognising cattle expressions if required.

23.2 INTRODUCTION

Qualitative Behaviour Assessment (QBA) is a method of animal welfare assessment developed in recent years at the Scottish Agricultural College (Wemelsfelder et al., 2000, 2001). This method relies on the ability of human observers to integrate perceived details of behaviour and its context into judgements of animal 'body language', using descriptors such as 'calm', 'tense', anxious' or 'content'. Such terms have previously been used to describe individual differences in animal temperament and personality (e.g. Lanier et al., 2000; Gosling, 2001), but they also have an expressive, emotional connotation that provides information which is directly relevant to animal welfare and could supplement information obtained from quantitative indicators (e.g. Carlstead et al., 1999; Weiss et al., 2006; Wemelsfelder and Farish, 2004; Wemelsfelder, 2007). Several years of research with pigs, cattle, poultry, sheep, and horses have shown QBA to have high inter- and intra-observer reliability and to be coherent with quantitative behavioural and physiological measures, both when animals were assessed individually and at group level (Wemelsfelder et al., 2000, 2001; Dungey, 2003; Wang, 2004; Rousing and Wemelsfelder, 2006; Napolitano et al., 2008; Minero et al., 2009).. In an on-farm context, an important question is whether qualitative judgements of behaviour might not be biased by the prevailing environmental conditions. In a study with individual pigs, digital manipulation of videos demonstrated that whether pigs were observed against an indoor or outdoor background did not unduly bias observers' assessment of the pigs' behaviour (Wemelsfelder et al., 2009).

These studies were all based on a Free-Choice-Profiling (FCP) methodology which asks observers to develop their own descriptive terminologies, and requires a minimum of 10 observers to be statistically viable. Evidently this method is unsuitable for on-farm welfare monitoring purposes, and requires adaptation for use by individual assessors. It was therefore the aim of the present study to develop, and test the inter-observer reliability of, a fixed qualitative rating scale describing the expressive behavioural repertoire ('body language') of cattle. More specifically, the aim was to investigate the reliability of such a scale for beef cattle, dairy cattle and veal calves kept under a variety of environmental conditions.

23.3 METHODS

23.3.1 Development of Rating Scale

A first selection of qualitative descriptors of cattle behaviour was made on the basis of several previous FCP-based studies of on-farm dairy cattle welfare (e.g. Rousing and Wemelsfelder, 2006). These studies, using different observer groups of cattle experts and animal welfare students, tended to find 2–3 main dimensions of cattle behavioural expression. These dimensions are characterised by clusters of terms at both their positive and negative ends; from these clusters, a list of 26 terms was selected as a starting-point for this study. This list was sent to 9 research partners in the Welfare Quality cattle-subgroup for comments and amendments, of which 6 replied. Feedback of these respondents was generally positive and various suggestions for additional descriptors were made. We decided it would be unwise to be too restrictive in the number of terms we selected at this early stage, and that a possible reduction of terms would be better based on the results of this study. Thus the adjusted list contained 29 descriptors for calves and dairy cattle, and 32 terms for beef cattle.

The rating scale to be tested was construed by putting each of the 29 (32) qualitative descriptors next to a continuous, undivided visual analogue scale of 125 mm length (the same as used in FCP studies). Each descriptor was given in English, and, for the on-farm work with veal calves and dairy cattle in Italy, an Italian translation in smaller font was also added to each term. No definitions of the terms were given, but observers were provided with detailed instructions on how to make qualitative assessments of the behaviour of animals in groups, and were encouraged to discuss and agree on the meaning of the terms as much as possible before starting the work.

23.3.2 Reliability Testing

Beef Cattle

Assessments were carried out on 21 groups of beef cattle distributed over 14 beef cattle farms located in central Scotland. The cattle groups differed with regard to breed, age, group size, and housing system. The breeds assessed in the present study included purebred Limousin, Charolais, Simmental, Salers, Belted Galloway, Luing, Beef Shorthorn and Highland cattle, and several commercial crossbreeds. A total of 21 cattle groups (of which 5 also contained calves) were observed in three different systems: indoor loose housing systems with slatted floors (5 groups); loose housing systems with deep straw bedding (9 groups); and free range outdoor systems (7 groups). The assessors were 3 graduate students and 1 undergraduate student (3 male, 1 female), three of whom had lived on or directly next to, a cattle farm, and two of whom had experience in measuring behaviour.

Dairy Cattle

Assessments were carried out on 22 groups of dairy cattle distributed over 17 dairy farms (6 tie stall, 6 cubicle and 5 straw-bedded farms), located in Southern Italy. The average number of lactating animals was 12, 80 and 60 for tie stall, cubicle and straw bedded farms respectively. All cows were Friesian. A total of 22 groups of lactating cows were observed (6 groups in tie stalls, 10 groups in cubicle systems, 6 groups in straw bedded pens). The observations were performed after morning milking (10.00–13.00). Assessors were 2 students of animal science (1 female, 1 male) and 2 researchers in the field of animal science (1 female, 1 male).

Veal Calves

Assessments were carried out on 22 groups of veal calves distributed over 3 farms located in Northern Italy. In 1 farm calves were housed in boxes containing about 30 animals, and milk was offered to these animals through an automatic milk dispenser placed in each box. On this farm 7 groups of calves were observed in the afternoon. In the other 2 farms calves were kept in pens each containing of 4–6 animals, and milk was distributed twice a day using a gun dispenser. On these farms 15 groups of calves were observed, either after milk distribution in the morning (7 groups), or immediately before milk distribution in the afternoon (8 groups). The calves were all Friesian, mainly originating in central Europe (Slovakia, Poland, France, Germany), and their age ranged from 45 to 180 days of life. The assessors were 2 State Veterinary Officers (1 female, 1 male), 1 practicing veterinarian (female), and 1 student of animal science (female).

Assessment Procedures (the Same for All Three Cattle Groups)

Before the first assessment, the four assessors recruited for each cattle group familiarised themselves with the rating scale and the written instructions. They discussed the assessment procedures as described in the instructions, but refrained from discussing the meaning of descriptors to enhance independence of assessment.

The assessments were then carried out in accordance with the instructions. One assessor was responsible for selecting suitable observation points and the timing of the observations. Assessors observed the same groups of calves/cattle at the same moment, standing closely together, but without distracting each other or blocking each other's view. Observation sessions lasted from 10 to 20 minutes, with the time spent at each observation point ranging from 2.5 (8 points) to 10 minutes (1 or 2 points). Having finished all observations, the assessors left the cattle unit and each scored the provided rating scale. They did not talk to each other during the entire procedure.

When all farm visits had been completed, the scores for each descriptor provided by the four assessors for each cattle group were recorded by measuring with a ruler the distance in millimetres between the minimum point on the left side of the scale and the mark on the line made by the assessor. The inter-observer reliability of these scores was calculated using Kendall Correlation Coefficient W. Subsequently a Principle Component Analysis (PCA; covariance matrix, no rotation) of scores was conducted for each individual assessor, to gain more insight in how each assessor had used the various terms to account for the variation between observed groups of calves/cattle. The scores attributed to the observed animal units on the first two main components of the PCA were then also tested for inter-observer reliability using Kendall W. The outcomes of the PCA then directed us to some further calculations and a reduction of the number of descriptors on the QBA rating scales, which are described below under 'Results'.

23.4 RESULTS

Table 23.1 shows the Kendall W values for each of the qualitative descriptors on the rating scales of the three cattle groups. Setting an arbitrary minimum of 0.7 (Martin and Bateson, 1993), 20 terms out of 32 (highlighted in black) reached satisfactory inter-observer reliability for beef cattle (rounded up from 0.65). Another 9 terms reached W values between 0.6 and 07 (rounded up from 0.55). For dairy cattle, 1 out of 29 terms reached satisfactory inter-observer reliability, and for veal calves 2 out of 29, with another 4 terms reaching W values between 0.6 and 0.7.

The assessors used their terms in very similar ways to distinguish between expressions of positive and negative mood (PC1), and between high and low levels of arousal in these

TABLE 23.1 Kendall correlation coefficients for all descriptors.

Descriptor	Kendall's W			Descriptor	Kendall's W		
	Beef cattle	Dairy cattle	Veal calves		Beef cattle	Dairy cattle	Veal calves
Welfare overall	.85***	.42*	.24	Distressed	.78***	.38	.31
Playful	.85***	.32	.48**	Frustrated	.75***	.36	.38
Active	.83***	.37	.65***	Agitated	.73***	.35	.50**
Boisterous	.80***	.32	.50**	Nervous	.68***	–	–
Content	.78***	.40*	.35	Uncomfortable	.68***	–	–
Lively	.77***	.44**	.65***	Vigilant/watchful	.66***	.42*	.51**
Friendly	.77***	.53**	.60***	Tense	.65***	.32	.38
Inquisitive	.75***	.65***	.59***	Uneasy	.64***	.36	.55***
Relaxed	.73***	.35	.38	Irritable	.64***	.36	.36
Sociable	.73***	.51	.54	Bored	.61**	.25	.26
Happy	.71***	.27	.39*	Scared	.61**	–	–
Calm	.69***	.39*	.52***	Fearful	.61**	.34	.57***
Positively occupied	.67***	.50**	.34	Indifferent	.61**	.19	.38
Confident	.63***	.37	.47**	Depressed	.58**	.36	.50**
Enjoying	.61**	.37	.32	Unwell	.53**	.38	.31
				Aggressive	.49*	.33	.27
				Apathetic	.23	.31	.56**

Notes: N = 4; df = 21: dairy, veal; 20: beef; *** p <0.001; ** p<0.01; * p<0.05.

moods (PC2). This similarity is also evident between the three cattle groups, even if there are slight differences in the terms characterising the PCA components (eg high loading terms on PC1: 'inquisitive' for calves, 'content' for beef cattle). For some assessors the meaning of their principal components appears to have 'swapped', so that PC2 distinguishes between positive and negative mood, and PC1 between levels of arousal. In these cases PC2 data have been amassed with the other assessors' PC1 data, and vice versa. Assessors for whom this was the case are marked in Tables 23.2, 23.3 and 23.4 with @.

Given this similarity, and the apparent face-value relevance of PC1 to welfare, we decided to make the integrated picture given by the PCA the focus for developing this measure, rather than the separate descriptors. We then tested the inter-observer reliability of the two main PCA components using Kendall W. These PCA-based outcomes are given in Tables 2 (beef cattle), 3 (dairy cattle), and 4 (veal calves).

To further prepare the rating scale for practical application, we reduced the number of descriptors of all groups to 20, by removing terms which 1) had relatively low loadings (lower or slightly higher than 0.20) on each of the assessors' PCA components, 2) had meanings without clear relevance to welfare, or 3) were similar in meaning to other terms on the list. We also took out 'welfare overall' as this term includes non-animal environmental features. For beef cattle this reduction resulted in the removal of 5 terms with Kendall W values of 0.7 or higher (welfare overall, playful, lively, agitated, vigilant), and 7 terms with W values lower than 0.7 (confident, scared, fearful, depressed, unwell, aggressive, apathetic). For dairy cattle and veal calves all terms removed had W values below 0.60 (welfare overall, confident, enjoying, vigilant, unwell, aggressive for both; boisterous, tense, depressed for dairy cattle; content, positively occupied, irritable for

TABLE 23.2 Beef cattle: PCA outcomes for full and reduced rating scales.

	PCA Factor 1 PCA of 28 terms / *PCA of 20 terms*					PCA Factor 2 PCA of 28 terms / *PCA of 20 terms*			
Kendall's W (N=4, df=21)	.75*** / *.73****					.85*** / *.84****			
% of variation explained	obs1 48 / *51*	obs2 42 / *48*	obs3 49 / *51*	obs4@ 26 / *26*		obs1 24 / *22*	obs2 33 / *30*	obs3 23 / *24*	obs4@ 39 / *41*
Main descriptors of PCA components	Content, calm, happy, relaxed, positively occupied, enjoying, friendly Uneasy, tense, uncomfortable, frustrated, distressed, nervous, irritable					Boisterous, inquisitive, sociable, active, friendly, enjoying Indifferent, calm, relaxed			
Correlation with 'welfare overall' (df=19)	obs1 .90***	obs2 .74***	obs3 .91***	obs4 .65***	mean .80***				

Notes: the main descriptors of PCA components are valid for both analyses; *** p <0.001; ** p<0.01; * p<0.05; @ assessor whose PC2 was treated as PC1 and vice versa (see text).

TABLE 23.3 Dairy cattle: PCA outcomes for full and reduced rating scales.

	PCA Factor 1 PCA of 28 terms / *PCA of 20 terms*					PCA Factor 2 PCA of 28 terms / *PCA of 20 terms*			
Kendall's W (N=4, df=21)	.42* / *.38*					.48** / *.46***			
% of variation explained	obs1 21 / *20*	obs2 48 / *49*	obs3 31 / *30*	obs4@ 25 / *27*		obs1 30 / *32*	obs2 15 / *19*	obs3 19 / *22*	obs4@ 45 / *45*
Main descriptors of PCA components	Content, inquisitive, sociable, positively occupied, happy, active, lively, friendly, playful, relaxed Bored, apathetic, frustrated, distressed					Irritable, uneasy, agitated, inquisitive, positively occupied, frustrated, fearful Indifferent, calm, relaxed, apathetic, bored, content			
Correlation with 'welfare overall' (df=19)	obs1 .74***	obs2 .88***	obs3 .44*	obs4 .24	mean .58				

Notes: the main descriptors of PCA components are valid for both analyses; *** p <0.001; ** p<0.01; * p<0.05; @ assessor whose PC2 was treated as PC1 and vice versa (see text).

TABLE 23.4 Veal calves: PCA outcomes for full and reduced rating scales.

	PCA Factor 1 PCA of 28 terms / *PCA of 20 terms*					PCA Factor 2 PCA of 28 terms / *PCA of 20 terms*			
Kendall's W (N=4, df=21)	.67*** / *.64****					.51** / *.40**			
% of variation explained	obs1 57 / *59*	obs2 46 / *47*	obs3 50 / *58*	obs4@ 20 / *20*		obs1 15 / *15*	obs2 15 / *17*	obs3 12 / *13*	obs4@ 37 / *41*
Main descriptors of PCA components	Inquisitive, sociable, friendly, playful, lively, happy, boisterous, active Depressed, bored, apathetic, frustrated, distressed, fearful, indifferent, tense, calm					Uneasy, tense, agitated, fearful, frustrated, distressed, boisterous, playful, lively, active, sociable, friendly, bored, apathetic Relaxed, calm, indifferent, happy			
Correlation with 'welfare overall' (df=19)	obs1 .72***	obs2 .46*	obs3 −.08	obs4 .36	mean .39				

Notes: the main descriptors of PCA components are valid for both analyses; *** p <0.001; ** p<0.01; * p<0.05; @ assessor whose PC2 was treated as PC1 and vice versa (see text).

calves). The remaining terms with low Kendall values stayed on the descriptor lists because of their high loadings on the PCA components. For beef cattle, of the 14 terms characterising PC1 (see Table 23.2) 11 had Kendall W values above 0.7. The reduced QBA rating scales for beef cattle, dairy cattle and veal calves thus consisted of the following 20 terms for all three categories: active, inquisitive, sociable, relaxed, calm, friendly, happy, frustrated, uneasy, bored, indifferent, distressed. In addition these scales also contained the terms, for beef cattle: positively occupied, boisterous, content, enjoying, uncomfortable, irritable, nervous, tense; for dairy cattle: positively occupied, playful, lively, content, fearful, agitated, irritable, apathetic; and for veal calves: playful, lively, boisterous, fearful, agitated, tense, depressed, apathetic.

We then checked the inter-observer reliability of the reduced rating scales by taking out the selected terms and their scores from the assessor files, conducting a PCA on these reduced files, and again calculating Kendall W for its main components. Finally, we investigated the extent to which assessors' PC1 scores reflected their view of the cattle/calves' overall welfare, by correlating these scores to their 'welfare overall' scores (Pearson or Spearman Rank correlations depending on the distribution of scores). These outcomes are also given in Tables 23.2, 23.3 and 23.4.

23.5 DISCUSSION

The aim of this study was to design a fixed rating scale of qualitative descriptors of cattle behavioural expression, and test its inter-observer reliability. In designing this scale we preferred to start with a rather extensive list of 29 terms (32 for beef cattle), and use the test results to remove less effective terms. Testing the rating scale with four assessors for each cattle group on 22(21) cattle units, we found for beef cattle that 20 out of 32 descriptors reached satisfactory reliability (similar to what was found for poultry and pigs). For dairy cattle and veal calves however, very few descriptors reached this criterion. It is not clear why this occurred – a subsequent video-based assessment of dairy cattle by 14 assessors using the same 20-term rating scale did find satisfactory reliability ($W=0.73$; $p<0.001$).

Despite the lack of agreement on separate descriptors for dairy cattle and calves, Principal Component Analysis (PCA) of assessor scores within each cattle group showed remarkably similar clusterings of descriptors along the two principal components (PCs). From these clusterings emerged a coherent pattern of cattle expression, in which the first PC distinguishes between positive and negative mood, and the second PC differentiates these moods in low and high levels of arousal. The second PC does not in and by itself appear to bear direct relevance to welfare, but contributes to a meaningful transition between descriptors of positive and negative mood on the first PC. Thus PC2 is an important part of the information on welfare provided by PC1; it is the overall expressive pattern that counts, and that provides farm units with their scores on PC1. The similarity of expressive patterns is also evident between the three cattle groups, even if there are slight differences in the terms characterising the PCA components.

These findings suggest that cattle welfare is reflected not so much by single terms, as by the pattern of behavioural expression that emerges from scoring a broader list of terms. This requires an integrative measure, which we propose can be provided by a farm unit's score on PC1. The PCA analysis for beef cattle appears sufficiently robust to support this proposal; PC1 (based on the 20-term PCA) showed satisfactory inter-observer reliability (0.73), and explained 26-51% of the variation between groups (depending on the assessor; 1 low value of 26% due to 'swapped' PCA components). Furthermore, PC1 showed a high mean correlation (0.80) to the assessors' score of 'welfare overall', which supports its direct relevance to beef cattle welfare. For dairy cattle and veal calves however PCA results were quantitatively weak. For dairy cattle, PC1 explained 20–49% of the variation and showed poor inter-observer reliability (0.38). This figure goes up to 0.56 however, if assessor 4 is taken out of the analysis. Assessor 4 also showed a low correlation of PC1 scores with 'welfare overall scores, whereas assessors 1 and 2 showed high correlations. For veal calves PC1 explained 20–59% of the variation, and showed reasonable inter-observer reliability (0.64), which goes up to 0.78 when assessor 4 is taken out of the analysis. Only assessor 1 showed a good correlation between PC1 and 'welfare overall' scores.

Thus our proposal that PC1 may provide an integrative measure of positive and negative mood mostly relies on our beef cattle results. However, our preference for this approach was strengthened by the results of the Welfare Quality QBA projects with poultry and pigs. In both these projects PCA revealed quantitatively robust patterns of poultry and pig expression that differentiated between positive and negative mood, and between low and high levels of arousal in these moods. In addition these patterns have emerged in a large number of comparable Free-Choice-Profiling studies at SAC. Thus, we think it is not unreasonable to assume that our PCA results reflect a pattern of animal behavioural expression that has wider validity, and that, given the beef cattle results, assessors should in principle be able to use reliably as an indicator of cattle welfare.

Clearly though, these results indicate that it is important to check the inter-observer reliability between assessors during the training period. This can be done from video, or, preferably, live. If problems arise, additional training in the observation and interpretation of cattle behavioural expression should be provided until a satisfactory consensus on the meaning of descriptors has been achieved.

Further testing and validation of a PCA-based measure can take place through investigation of its relationship to other welfare indicators, such as those that will be part of the WQ monitoring tool. Perhaps the proposal to make QBA monitoring dependent on multivariate analysis through PCA complicates the use of this measure as part of a larger monitoring tool; however, we think this complication will be offset by the advantages of having an integrative measure that contributes to the transparency and face-validity of that tool for the many stakeholder groups involved.

23.6 CONCLUSIONS

In conclusion, given QBA's satisfactory inter-observer reliability for beef cattle, its high relevance to cattle welfare, and its high feasibility, we suggest cattle QBA should be part of the WQ monitoring tool.

BIBLIOGRAPHY

Alban, L., Agger, J.F., Lawson L.G. (1996) 'Lameness in tied Danish dairy cattle: the possible influence of housing systems, management, milk yield and prior incidents of lameness', *Preventive Veterinary Medicine*, 29, pp. 135–149.

Albright, J.L. and Arave, C.W. (1997) 'Aggressive and submissive behaviour', in: J.L. Albright and C.W. Arave (eds) *The Behaviour of Cattle*, CAB International, Wallingford, pp. 59–66.

Algers, B. and Atkinson, S. (2006) 'Stun quality for cattle and pigs in relation to handling at slaughter', *Proceedings of the International Conference on Animal Health and Welfare of Food Producing Animals, 29–30 March, Athens*, p. 11.

Altman, D.G. (1991) *Practical Statistics for Medical Research*, 1st edn. Chapman & Hall. London.

Andersen, I.L., Bo, K.E., Foerevik, G., Janczak, A.M. and Bakken, M. (2000) 'Behavioural evaluation of methods for assessing fear responses in weaned pigs', *Applied Animal Behaviour Science*, 69, pp. 227–240.

Anderson, B. and Horder, J.C. (1979) 'The Australian carcass bruise scoring system', *Queensland Agricultural Journal*, 105, pp. 281–287.

Angus, L.J., Bowen, H., Gill, L.A.S., Knowles, T.G. and Butterworth, A. (2005) 'The use of conjoint analysis to determine the importance of factors that affect on-farm welfare of the dairy cow', *Animal Welfare*, 14(3), pp. 203–213.

Anon. (2001) 'Scientists' assessment of the impact of housing and management on animal welfare', *Journal of Applied Animal Welfare Science*, 4, pp. 3–52.

Apple, J.K., Kegley, E.B., Galloway, D.L., Wistuba, T.J. and Rakes, L.K. (2005) 'Duration of restraint and isolation stress as a model to study the dark-cutting condition in cattle', *Journal of Animal Science*, 83, pp. 1202–1214.

Askar, R. (2006) Veterinarian, Swedish National Food Administration, personal communication, 5 March.

Atkinson, S. (2000) *Farm Animal Transport, Welfare and Meat Quality*. MSc thesis, Swedish University of Agricultural Sciences.

Aureli, F., Preston, S.D. and de Waal, F.B. (1999) 'Heart rate response to social interactions in free-moving rhesus macaques (Macaca mulatta): a pilot study', *Journal of Comparative Physiology*, 113, pp. 59–65.

Bahrs, E. (2005) *Verhalten und Gesundheitsstatus von Mastbullen auf Gummispaltenboden*. Dissertation, Ludwig-Maximillian-Universität München.

Baldock, N.M., Sibly, R.M. and Penning, P.D. (1988) 'Behaviour and seasonal variation in heart rate in domestic sheep (Ovis aries)', *Animal Behaviour*, 36, pp. 35–43.

Bareille, N., Beaudeau, F., Billon S., Robert, A. and Faverdin, P. (2003) 'Effects of health disorders on feed intake and milk production in dairy cows.', *Livestock Production Science*, 83, pp. 53–62.

Bartussek, H., Leeb, C. and Held, S. (2000) *Animal Needs Index for Cattle*, Sheet 5, Stockmanship, ANI35L/2000. BAL Gumpenstein, Irdning.

Bergsten, C. and Pettersen, B. (1992) 'The cleanliness of cows tied in stalls and the health of their hooves as influenced by the use of electronic trainers', *Preventive Veterinary Medicine*, 13, pp. 229–238.

Berlyne, D.E. (1960) *Conflict, Arousal and Curiosity*. Mc Graw Hill, New York.

Bielfeldt, J.C., Badertscher, R., Tolle, K.H. and Krieter, J. (2005) 'Risk factors influencing lameness and claw disorders in dairy cows', *Livestock Production Science*, 95(3), pp. 265–271.

Blom, J.Y. (1983) 'Traumatic injuries and foot diseases as related to housing systems', in: S.H. Baxter, M.R. Baxter and J.A.D. MacCormack (eds) *Farm Animal Housing and Welfare*, Martinus Nijhoff Publishers, Boston, MA, pp. 216–225.

Boccia, M.L., Reite, M. and Laudenslager, M. (1989) 'On the physiology of grooming in a pigtail macaque', *Physiology and Behaviour*, 45, pp. 667–670.

Bøe, K.E. and Færevik, G. (2003) 'Grouping and social preferences in calves, heifers and cows', *Applied Animal Behaviour Science*, 80, pp. 175–190.

Boissy, A. (1995) 'Fear and fearfulness in animals', *Quarterly Reviews of Biology*, 70, pp. 165–191.

Boissy, A. and Bouissou, M.F. (1995) 'Assessement of individual differences in behavioural reactions of heifers exposed to various fear-eliciting situations', *Applied Animal Behaviour Science*, 46, pp. 17–31.

Boissy, A. and Dumont, B. (2002) 'Interactions between social and feeding motivations on the grazing behaviour of herbivores: sheep more easily split into subgroups with familiar peers', *Applied Animal Behaviour Science*, 79, pp. 233–245.

Boissy, A. and Le Neindre, P. (1990) 'Social influences on the reactivity of heifers: Implications for learning abilities in operant conditioning', *Applied Animal Behaviour Science*, 25, pp. 149–165.

Boivin, X., Lensink, J., Tallet, C. and Veissier, I. (2003) 'Stockmanship and farm animal welfare', *Animal Welfare*, 12, pp. 479–492.

Bokkers, E.A.M. and Koene, P. (2001) 'Activity, oral behaviour and slaughter data as welfare indicators in veal calves: a comparison of three housing systems', *Applied Animal Behaviour Science*, 75, pp. 1–15.

Boleman, S.L., Boleman, S.J., Morgan, W.W., Hale, D.S., Griffin, D.B., Savell, J.W., Ames, R.P., Smith, M.T., Tatum, J.D., Field, T.G., Smith, G.C., Gardner, B.A., Morgan, J.B., Northcutt, S.L., Dolezal, H.G., Gill, D.R. and Ray, F.K. (1998) 'National beef quality audit-1995: Survey of producer-related defects and carcass quality and quantity attributes', *Journal of Animal Science*, 76, pp. 96–103.

Bouissou, M.F. (1985) *Contribution à l'étude des relations interindividuelles chez les bovins dometiques femelles (Bos taurus L.)*. Thèse de Doctorat d'Etat, Université Paris VI.

Bowell, V.A., Rennie, L.J., Tierney, G., Lawrence, A.B. and Haskell, M.J. (2003) 'Relationships between building design, management system and dairy cow welfare', *Animal Welfare*, 12(4), pp. 547–552.

Bracke, M.B.M., Metz, J.H.M., Dijkhuizen, A.A. and Spruijt, B.M. (2001) 'Development of a decision support system for assessing farm animal welfare in relation to husbandry systems: strategy and prototype', *Journal of Agricultural and Environmental Ethics*, 14, pp. 321–337.

Brattberg, M. (2006) Veterinarian, Swedish National Food Administration, personal communication, 5 March.

Brenninkmeyer, C., Dippel, S., March S., Brinkmann, J., Winckler, C. and Knierim, U. (2007) 'Reliability of a subjective lameness scoring system for dairy cows', *Animal Welfare*, 16, pp. 127–129.

Breuer, K., Hemsworth, P.H., Barnett, J.L., Matthews, L.R. and Coleman, G.J. (2000) 'Behavioural response to humans and the productivity of commercial dairy cows', *Applied Animal Behaviour Science*, 66, pp. 273–288.

Broom, D.M. (1999) 'Animal welfare: the concept and the issues', in: F. L. Dolins (ed.) *Attitudes to Animals: News in Animal Welfare*, Cambridge University Press, Cambridge, pp. 129–142.

Broring, N., Wilton, J.W. and Colucci, P.E. (2003) 'Body condition score and its relationship to ultrasound backfat measurements in beef cows', *Canadian Journal of Animal Science*, 83, pp. 593–596.

Burrow, H.M. (1997) 'Measurement of temperament and their relationships with performance traits of beef cattle', *Animal Breeding Abstracts*, 65(7), pp. 477–495.

Busato, A., Trachsel, P. and Blum, J. (2000) F'requency of traumatic cow injuries in relation to housing systems in Swiss organic dairy herds', *Journal of Veterinary Medicine A*, 47, pp. 221–229.

Butler, W.R. (2003) 'Energy balance relationships with follicular development, ovulation and fertility in postpartum dairy cows', *Livestock Production Science*, 83, pp. 211–218.

Capdeville, J. and Veissier, I. (2001) 'A method of assessing welfare in loose housed dairy cows at farm level, focusing on animal observations', *Acta Agriculturae Scandinavica Section A: Animal Science*, 30(Suppl.), pp. 62–68.

Carlstead, K., Fraser, J., Bennet, C. and Kleiman, D.G. (1999) 'Black rhinoceros (Diceros bicornis) in US zoos: II. Behavior, breeding success, and mortality in relation to housing facilities', *Zoo Biology*, 18(1), pp. 35–52.

Chaplin, S.J., Tierney, G., Stockwell, C., Logue, D.N. and Kelly, M. (2000) 'An evaluation of mattresses and mats in two dairy units', *Applied Animal Behaviour Science*, 66, pp. 263–272.

Clarkson, M.J., Downham, D.Y., Faaull, W.B., Hughes, J.W., Manson, F.J., Merritt., J.B., Murray, R.D., Russell, W.B., Sutherst, J.E. and Ward, W.R. (1996) 'Incidence and prevalence of lameness in dairy cattle', *Veterinary Record*, 138, pp. 563–567.

Cochran, W.G. (1977) *Sampling Techniques*, 3rd edn. Wiley & Sons, New York.

Coles, G.C., Hadley, P.J., Milnes, A.S., Green, L.E., Stosic, P.J. and Garnsworthy, P.C. (2003) 'Relationship between lice infestation and leather damage in cattle', *Veterinary Record*, 152, pp. 255–259.

Collier, R.J., Dahl, G.E., VanBaale, M.J. (2006) 'Major advances associated with environmental effects on dairy cattle', *Journal of Dairy Science*, 89(4), pp. 1244–1253.

Colwell, D.D. and Himsl-Rayner, C. (2002) 'Linognathus vituli (Anoplura : Linognathidae): population growth, dispersal and development of humoral immune responses in naive calves following induced infestations', *Veterinary Parasitology*, 108, pp. 237–246.

Cook, N.B., Bennett, T.B. and Nordlund, K.V. (2005) 'Monitoring indices of cow comfort in free-stall-housed dairy herds', *Journal of Dairy Science*, 88(11), pp. 3876–3885.

Cook, N. (n.d.) *Clinical Information and Forms: Hygiene*, published online <http://www.vetmed.wisc.edu/dms/fapm/fapmtools/hygiene.htm>.

Coppedge, B.R., Carter, T.S., Shaw, J.H. and Hamilton, R.G. (1997) 'Agonistic behaviour associated with orphan bison (Bison bison L.) calves released into a mixed resident population', *Applied Animal Behaviour Science*, 55, pp. 1–10.

Cozzi, G., Ricci, R., Dorigo, M. and Zanet, D. (2005) 'Growth performance, cleanliness and lameness of finishing Charolais bulls housed in littered pens of different design', *Italian Journal of Animal Science*, 4(Suppl. 2), pp. 251–253.

Croxton, D. and Stollard, R.J. (1976) 'Use of body condition scoring as a management aid in dairy and beef herds' (abstract), *Animal Production*, 22, pp. 146–147.

Cullor, J.S. (1997) 'HACCP: is it coming to the dairy?', *Journal of Dairy Science*, 80, pp. 3449–3452.

Danuser, J. and Regula, G. (2001) *Evaluation der Ökomassnahmen und Tierhaltungsprogramme: Bereich Artgerechte Tierhaltung*, Vierter Zwischenbericht. Bundesamt für Veterinärwesen, Bern.

de Passillé, A.M., Rushen, J. and Martin, F. (1995) 'Interpreting the behaviour of calves in an open-field test: a factor analyses', *Applied Animal Behaviour Science*, 45, pp. 201–213.

de Passillé, A., Rushen, J., Ladewig, J. and Petherick, C. (1996) 'Dairy calves discrimination of people based on previous handling', *Journal of Animal Science*, 74, pp. 969–974.

De Rosa, G., Tripaldi, C., Napolitano, F., Saltalamacchia, F., Grasso, F., Bisegna, V. and Bordi, A. (2003) 'Repeatability of some animal-related variables in dairy cows and buffoloes', *Animal Welfare*, 12, pp. 625–629.

de Wilt, J.G. (1985) *Behaviour and Welfare of Veal Calves in Relation to Husbandry Systems*. PhD thesis, Landbouwhogeschool Wageningen.

Dellmeier, G.R., Friend, T.H. and Gbur, E.E. (1985) 'Comparison of four methods of calf confinement: II. Behaviour', *Journal of Animal Science*, 60, pp. 1102–1109.

Dellmeier, G., Friend, T. and Gbur, E. (1990) 'Effects of changing housing on open-field behaviour of calves', *Applied Animal Behaviour Science*, 26, pp. 215–230.

Désiré, L., Boissy, A. and Veissier, I. (2002) 'Emotions in farm animals: a new approach to animal welfare in applied ethology', *Behavioural Processes*, 60, pp. 165–180.

D.L.G. (Deutsche Landwirtschaftsgesellschaft) (2000) *Prüfprogramm und Beurteilungsschema für die DLG-Prüfung von elastischen Stallbelägen*. D.L.G, Frankfurt am Main.

Dungey, T.M. (2003) *The Quantitative and Qualitative Assessments of Behavioural Reactions of Sheep (Ovis aries) towards Fear-eliciting Situations: Effects of Breed and Physiological State*. MSc dissertation, University of Edinburgh.

Dyer, R.M., Rajkondawar, P.G., Neerchal, N.K., Tasch, U., Lefcourt, A.M. and Varner, M.A. (2004) 'The role of lesion and gait analyses in models of bovine lameness', in: B. Zemljic (ed.) *Proceedings of the 13th International Symposium and 5th Conference on Lameness in Ruminants*, Maribor, pp. 76–77.

Earle, D.F. (1976) 'A guide to scoring dairy cow condition', *Journal of the Department of Agriculture Victoria*, 74, p. 228.

Edmondson, A.J., Lean, I.J., Weaver, L.D., Farver, T. and Webster, G. (1989) 'A body condition scoring chart for Holstein dairy cows', *Journal of Dairy Science*, 72, pp. 68–78.

EFSA (2006) 'Scientific report on the risks of poor welfare in intensive calf farming systems: an update of the scientific veterinary committee report on the welfare of calves', *EFSA Journal*, 366, pp. 1–36.

Eikelenboom, G. (ed.) (1983) *Stunning of Animals for Slaughter*. Martinus-Nijhoff, The Hague.

Ekesbo, I. (1966) 'Disease incidence in tied and loose housed dairy cattle', *Acta Agriculturae Scandinavica*, 15(Suppl.), p. 74.

Emmerig, H. (2004) *Behavioural Indicators of Good Welfare in Dairy Cows: An Exploratory Approach*. Diploma Thesis, University of Kassel.

Engel, B. and van Reenen, K. (2006) *Intra and Inter-observer Variability: Note for EC Welfare Project WP 2.1*. Wageningen University.

Erhard, H.W. and Schouten, J.M. (2001) 'Individual differences and personality', in: L.J. Keeling and H.W. Gonyou (eds) *Social Behaviour in Farm Animals*, CAB International, Wallingford, pp. 333–352.

ESCA (1976) *Condition Scoring: Dairy Cows*, Advisory Leaflet no. 100. East of Scotland College of Agriculture, Ayr.

Evans, D.G. (1978) 'The interpretation and scoring of subjective body condition scores', *Animal Production*, 26, pp. 119–125.

Fabianson, F., Erichsen, I., Reuterswärd, A.L. and Malmfors, G. (1982) 'The incidence of dark cutting beef in Sweden', *Meat Science*, 10, pp. 21–33.

Fagen, R.M. (1981) *Animal Play Behaviour*. Oxford University Press, New York.

Faye, B. and Barnouin, J. (1985) Objectivation de la propreté des vaches laitières et des stabulations. L'indice de propreté. *Bulletin Technique C.R.Z.V. de Theix I.N.R.A.*, 59, pp. 61–67.

Feh, C. and de Mazieres, J. (1993) 'Grooming at a preferred site reduces heart rate in horses', *Animal Behaviour*, 46, pp. 1191–1194.

Ferguson, J.D., Galligan, D.T. and Thomsen, N. (1994) 'Principal descriptors of body condition score in Holstein cows', *Journal of Dairy Science*, 77, pp. 2695–2703.

Fisher, A.D., Morris, C.A. and Matthews, L.R. (2000) 'Cattle behaviour: comparison of measures of temperament in beef cattle', *Proceedings of the New Zealand Society of Animal Production*, 60, pp. 214–217.

Fleiss, J.L. (1971) 'Measuring nominal scale agreement among many raters', *Psychological Bulletin*, 76, pp. 378–382.

Flower, F.C. (2006) 'Effects of milking on dairy cow gait', *Journal of Dairy Science*, 89, pp. 2084–2089.

Food Standards Agency (2002) *Red Meat Safety and Clean Livestock*. Food Standards Agency, London.

Forkman, B., Boissy, A., Meunier-Salaün, M.-C., Canali, E. and Jones, R.B. (2007) 'A critical review of fear tests used on cattle, pigs, sheep, poultry and horses', *Physiology and Behavior*, 92, pp. 340–374.

Forkman, B., Furuhaug, I.L. and Jensen, P. (1995) 'Personality, coping patterns, and aggression in piglets', *Applied Animal Behaviour Science*, 45, pp. 31–42.

Fowler, H. (1965) *Curiosity and Exploratory Behaviour*. Macmillan, New York.

Fraser, A.F. and Broom, D.M. (1990) *Farm Animal Behaviour and Welfare*, 3rd edn. Baillière Tindall, London

Fraser, A.F. and Broom, D.M. (1997) *Farm Animal Behaviour and Welfare*. CAB International, Wallingford.

Fregonesi, J. and Leaver, J.D. (2001) 'Behaviour, performance and health indicators of welfare for dairy cows housed in strawyard or cubicle systems', *Livestock Production Science*, 68, pp. 205–216.

Friedli, K., Schaub, J. and Wechsler, B. (1999) 'Strohmatratzen und weiche Liegematten für Milchvieh-Boxenlaufställe im Vergleich', *14. IGN-Tagung und 6. Freiland-Tagung*, pp. 42–45.

Gearhart, M.A., Curtis, C.R., Erb, H.N., Smith, R.D., Sniffen, C.J., Chase, L.E. and Cooper, M.D. (1990) 'Relationship of changes in condition score to cow health in Holsteins', *Journal of Dairy Science*, 73, pp. 3132–3140.

Gosling, S.D. (2001) 'From mice to men: what can we learn about personality from animal research?', *Psychological Bulletin*, 127, pp. 45–86.

Gottardo, F., Fregolent, G., Ravarotto, L. and Cozzi, G. (2003) 'Welfare and meat quality of beef cattle housed on two types of floors with the same space allowance', Italian *Journal of Animal Science*, 2(4), pp. 243–253.

Gottardo, F., Ricci, R., Preciso, S., Ravarotto, L. and Cozzi, G. (2004) 'Effect of the manger space on welfare and meat quality of beef cattle', *Livestock Production Science*, 89(2/3), pp. 277–285.

Graat, E.A.M., Frankena, K. and Bos, H. (1997) 'Principles and methods of sampling in animal disease surveys', in: J.P.T.M. Noordhuizen, K. Frankena, M.V. Thrusfield and E.A.M. Graat (eds) *Application of Quantitative Methods in Veterinary Epidemiology*, 2nd edn, Wageningen Pers, Wageningen.

Graf, B. (1994) 'Abnormal oral behaviours in fattening bulls: incidence, causation and implications', *Applied Animal Behaviour Science*, 40, pp. 79–80.

Grainger, C. and McGowan, A.A. (1982) 'The significance of pre-calving nutrition of the dairy cow', *New Zealand Society of Animal Procustion Occasional Publications*, 8, pp. 134–171.

Grandin, T. (1996) 'Animals welfare in slaughter plants', *Proceedings of the 29th Annual Conference of the American Association of Bovine Practitioners*, pp. 22–26.

Grandin, T. (1997) 'Assessement of stress during handling and transport', *Journal of Animal Science*, 75, pp. 249–257.

Grandin, T. (1998) 'The feasibility of using vocalisation scoring as an indicator of poor welfare during cattle slaughter', *Applied Animal Behaviours Science*, 56, pp. 121–128.

Grandin, T. (2005) 'Maintenance of good animal welfare standards in beef slaughter plants by use of auditing programs', *Journal of the American Veterinary Medical Association*, 226, pp. 370–373.

Grandin, T. (2005) *Recommended Animal Handling Guidelines and Audit Guide*, 2005 edn. AMI Foundation, Washington, DC.

Gregory, N.G. (1998) *Animal Welfare and Meat Science*. CABI Publishing, Wallingford.

Grignard, L., Boissy, A., Boivin, X., Garel, J.P. and Le Neindre, P. (2000) 'The social environment influences the behavioural responses of beef cattle to handling', *Applied Animal Behaviour Science*, 68, pp. 1–11.

Grignard, L., Boivin, X., Boissy, A. and Le Neindre, P. (2001) 'Do beef cattle react consistently to different handling situations?', *Applied Animal Behaviour Science*, 71, pp. 263–276.

Groth, W. (1984) 'Mängel im Haltungssystem als Ursache von Gliedmassenschäden bei Rindern', *Tierärztliche Umschau*, 39, pp. 196–201.

Groth, W. (1985) 'Kriterien für die Beurteilung von Haltungssystemen für Milchkühe und Mastbullen aus klinischer Sicht', *Tierärztliche Umschau*, 40, pp. 739–750.

Gruner, J. (1972) 'Einfluss der Fussbodengestaltung auf die Klauen- und Gliedmassengesundheit bei Mastbullen', *Monatshefte der Veterinärmedizin*, 27, pp. 273–276.

Gustafson, G.M. (1993) 'Effects of daily exercise on the health of tied dairy cows', *Preventive Veterinary Medicine*, 17, pp. 209–223.

Gygax, L., Westerath, H.S., Kuhlicke, J., Wechsler, B. and Mayer, C. (2005) 'Assessing cubicle dimensions for finishing bulls based on animal behaviour and cleanliness', *Animal Science*, 81(3), pp. 423–430.

Haeger, W. (2006) Veterinarian, Swedish National Food Administration, personal message, 15 June.

Hansen, K., Strom, J.S. and Levring, M. (2003) 'Force measurements in welfare evaluation of cattle housing design', *Applied Engineering in Agriculture*, 19(2), pp. 227–233.

Haskell, M.J., Rennie, L.J., Bowell, V.A., Wemelsfelder, F. and Lawrence, A.B. (2003) 'On-farm assessment of the effect of management and housing type on behaviour and welfare in dairy cattle', *Animal Welfare*, 12(4), pp. 553–556.

Hassall, S.A., Ward, W.R. and Murray, R.D. (1993) 'The effect of lameness on the behaviour of cows during summer', *Veterinary Record*, 132, pp. 578–580.

Heeres, J.J., Ruis-Heutinck, L.F.M. and van Reenen, C.G. (2000) 'The effect of roughage supply on zootechnical parameters, meat quality, behaviour and pathological status of the digestive organs in veal calves', in: J.A.M. Van Arendonk, A. Hofer, Y. van der Honing, F. Madec, K. Sejrsen, D. Pullar, L. Bodin, J.A. Fernández and E.W. Bruns (eds) *51st Annual Meeting of EAAP*, The Hague, p. 199.

Hemsworth, P.H. (2003) 'Human–animal interactions in livestock production', *Applied Animal Behaviour Science*, 81, pp. 185–198.

Hemsworth, P.H. and Coleman, G.J. (1998) *Human–Livestock Interactions: The Stockperson and the Productivity of Intensively Farmed Animals*. CAB International: Wallingford.

Hemsworth, P.H., Price, E.O. and Borgwardt, R. (1996) 'Behavioural responses of domestic pigs and cattle to humans and novel stimuli', *Applied Animal Behaviour Science*, 50, pp. 43–56.

Herskin, M.S., Kristensen, A.-M. and Munksgaard, L. (2004) 'Behavioural responses of dairy cows toward novel stimuli presented in the home environment', *Applied Animal Behaviour Science*, 89, pp. 27–40.

Hickey, M.C., Earley, B., Fisher, A.D. (2003) 'The effect of floor type and space allowance on welfare indicators of finishing steers', *Irish Journal of Agricultural and Food Research*, 42(1), pp. 89–100.

Hindhede, J., Sørensen, J.T., Jensen, M.B. and Krohn, C.C. (1996) 'Effect of space allowance, access to bedding, and flock size in slatted floor systems on the production and health of dairy heifers', *Acta Agriculturae Scandinavica Section A: Animal Science*, 46, pp. 46–53.

Hindhede, J., Mogensen, L. and Sørensen, J.T. (1999) 'Effect of group composition and feeding system on behaviour, production and health of dairy heifers in deep bedding systems', *Acta Agriculturae Scandinavica Section A: Animal Science*, 49, pp. 211–220.

Horning, B. (2003) 'Attempts to integrate different parameters into an overall picture of animal welfare using investigations in dairy loose houses as an example', *Animal Welfare*, 12(4), pp. 557–563.

Houghton, P.L., Lemenager, P.L., Moss, G.E. and Hendrix, K.S. (1990) 'Prediction of post-partum beef cow body composition using weight to height ratio and body condition score', *Journal of Dairy Science*, 68, pp. 1428–1437.

Hughes J. (2001) 'A system for assessing cow cleanliness', *In Practice*, 23, pp. 517–524.

Hughes, B.O. and Curtis, P.E. (1997) 'Health and disease', in: M.C. Appleby and B.O. Hughes (eds), *Animal Welfare*, CAB International, Wallingford, pp. 109–125.

Huxley, J. and Whay, H.R. (2006) 'Cow based assessments: Part 1. Nutrition, cleanliness and coat condition', *UK Vet*, 11(3), pp. 18–23.

Jarvis, A.M., Selkirk, L. and Cockram, M.S. (1995) 'The influence of source, sex class and pre-slaughter handling on the bruising of cattle at two slaughterhouses', *Livestock Production Science*, 43, pp. 215–224.

Jefferies, B.C. (1961) 'Body condition scoring and its use in management', *Tasmanian Journal of Agriculture*, 32, p. 19.

Jensen, M.B. (1999) 'Effects of confinement on rebounds of locomotor behaviour of calves and heifers, and the spatial preferences of calves', *Applied Animal Behaviour Science*, 62, pp. 43–56.

Jensen, M.B. and Kyhn, R. (2000) 'Play behaviour in group-housed dairy calves, the effect of space allowance', *Applied Animal Behaviour Science*, 67, p. 35.

Jensen, M.B., Vestergaard, K.S. and Krohn, C.C. (1998) 'Play behaviour in dairy calves kept in pens: the effect of social contact and space allowance', *Applied Animal Behaviour Science*, 56, p. 97.

Jonson, K. (2006) Swedish Meats, personal communication, 30 March.

Jonsson, K. (2006) Swedish Meats, personal communication, 22 March.

Keil, N.M., Wiederkehr, T.U., Friedli, K. and Wechsler, B. (2006) 'Effects of frequency and duration of outdoor exercise on the prevalence of hock lesions in tied Swiss dairy cows', *Preventive Veterinary Medicine*, 74, pp. 142–153.

Kendall, P.E., Nielsen, P.P., Webster, J.R., Verkerk, G.A., Littlejohn, R.P. and Matthews, L.R. (2006) 'The effects of providing shade to lactating dairy cows in a temperate climate', *Livestock Science*, 103(1/2), pp. 148–157.

Keverne, E.B., Martensz, N.D. and Tuite, B. (1989) 'Beta-endorphin concentrations in cerebrospinal fluid of monkeys are influenced by grooming relationships', *Psychoneuroendocrinology*, 14, pp. 155–161.

Kilgour, R.J., Melville, G.J. and Greenwood, P.L. (2006) 'Individual differences in the reaction of beef cattle to situations involving social isolation, close proximity of humans, restraint and novelty', *Applied Animal Behaviour Science*, 99, pp. 21–40.

Klaas, I.C., Rousing, T., Fossing, C., Hindhede, J. and Sørensen, J.T. (2003) 'Is lameness a welfare problem in dairy farms with automatic milking systems?', *Animal Welfare*, 12, pp. 599–603.

Knierim, U. (1999) 'Das Verhalten von Färsen bei der Einzel- oder Gruppeneinführung in die Milchviehherde', *KTBL-Schrift*, 382, pp. 115–120.

Knierim, U., Carter, C.S., Fraser, D., Gärtner, K., Lutgendorf, S.K., Mineka, S., Panksepp, J. and Sachser, N. (2001) 'Good welfare: improving quality of life', in: D.M. Broom (ed.) *Coping with Challenge: Welfare in Animals Including Humans*, Dahlem Workshop Report, Dahlem University Press, Berlin, pp. 79–100.

Kögler, H. (2003) *Schäden am Integument: Einfluss von Einstreu auf die Gelenkgesundheit bei Milchkühen*. Diploma Thesis, Technische Universität München.

Kondo, S., Sekine, J., Okubo, M. and Asahida, Y. (1989) 'The effect of group size and space allowance on the agonistic and spacing behavior of cattle', *Applied Animal Behaviour Science*, 24, pp. 127–135.

Kött & Charkföretagen (2006) *Information från KCF 1: Ny norm för bedömning av gödselförorenade slaktdjur*. KCF, Stockholm.

Krebs, S., Danuser, J. and Regula, G. (2001) 'Using a herd monitoring system in the assessment of welfare', *Acta Agricultura Scandinavica A*, 30, pp. 78–81.

Krohn, C.C. (1994) 'Behaviour of dairy cows kept in extensive (loose housing/pasture) or intensive (tie stall) environments: III. Grooming, exploration and abnormal behaviour', *Applied Animal Behaviour Science*, 42(2), pp. 73–86.

Kümper, H. (2003) 'Abszesse und Schleimbeutelentzündungen in einem Milchrinderbestand', *Tierärztliche Praxis*, 31G, pp. 306–307.

Lanier, J.L., Grandin, T., Green, R.D., Avery, D. and Mcgee, K. (2000) 'The relationship between reaction to sudden, intermittent movements and sounds and temperament', *Journal of Animal Science*, 78, pp. 1467–1474.

Lawrence, A. (1987) 'Consumer demand theory and the assessment of animal welfare', *Animal Behaviour*, 35, pp. 293–295.

Leach, K.A., Dippel, S., Huber, J., March, S., Winckler, C. and Whay, H.R. (2009) 'Assessing lameness in cows kept in tie-stalls', *Journal of Dairy Science*, 92, pp. 1575–1580.

Leeb, C., Main, D.C.J., Whay, H.R. and Webster, A.J.F. (2004) *Bristol Welfare Assurance Programme: Cattle Assessment*. University of Bristol.

Lensink, B.J., Fernandez, X., Boivin, X., Pradel, P., Le Neindre, P. and Veissier, I. (2000) 'The impact of gentle contacts on ease of handling, welfare, and growth of calves, and quality of veal meat', *Journal of Animal Science*, 78, pp. 1219–1226.

Lensink, B.J., Boivin, X., Pradel, P., Le Neindre, P. and Veissier, I. (2000a) 'Reducing veal calves' reactivity to people by providing additional human contact', *Journal of Animal Science*, 78, pp. 1213–1218.

Lensink, J.B., Boissy, A. and Veissier, I. (2000b) 'The relationship between farmers' attitude and behaviour toward calves, and productivity of veal units', *Annales de Zootechnie*, 49, pp. 313–327.

Lensink, B.J., Fernandez, X., Cozzi, G., Florand, L. and Veissier, I. (2001a) 'The influence of farmers' behavior on calves' reactions to transport and quality of veal meat', *Journal of Animal Science*, 79, pp. 642–652.

Lensink, B.J., Raussi, S., Boivin, X., Pykkonen, M. and Veissier, I. (2001b) 'Reactions of calves to handling depend on housing condition and previous experience with humans', *Applied Animal Behaviour Science*, 70, pp. 187–199.

Lensink, B., Reenen, C.V., Engel, B., Rodenburg, T. and Veissier, I. (2003) 'Repeatability and reliability of an approach test to determine calves' responsiveness to humans: "a brief report"', *Applied Animal Behaviour Science*, 83, pp. 325–330.

Lidfors, L.M. (1993) 'Cross-sucking in group-housed dairy calves before and after weaning off milk', *Applied Animal Behaviour Science*, 38, pp. 15–24.

Livesey, C.T., Marsh, C., Metcalf, J.A. and Laven, R.A. (2002) 'Hock injuries in cattle kept in straw yards or cubicles with rubber mats or mattresses', *Veterinary Record*, 150, pp. 677–679.

Loberg, J., Telezhenko, E., Bergsten, C. and Lidfors, L. (2004) 'Behaviour and claw health in tied dairy cows with varying access to exercise in an outdoor paddock', *Applied Animal Behaviour Science*, 89(1/2), pp. 1–16.

Lowman, B.G., Scott, N.A. and Somerville, S.H. (1973) *Condition Scoring of Cattle*, Bulletin no. 6. ESCA Animal Production Advisory Department.

MAFF (1986) Condition Scoring of Dairy Cows, Leaflet 612, MAFF, London.

Main, D.C.J., Whay, H.R., Green, L.E. and Webster, A.J.F. (2003) 'Preliminary investigation into the use of expert opinion to compare the overall welfare of dairy cattle farms in different farm assurance schemes', *Animal Welfare*, 12(4), pp. 565–569.

Main, D.C.J., Kent, J.P., Wemelsfelder, F., Ofner, E. and Tuyttens, F.A.M. (2003) 'Application of methods of on-farm welfare assessment', *Animal Welfare*, 12(4), pp. 573–578.

Manson, F.J. and Leaver, J.D. (1988) 'The influence of concentrate amount on locomotion and clinical lameness in dairy cattle,' *Animal Production*, 47, pp. 185–190.

March, S., Brinkmann, J. and Winckler, C. (2007) 'Effect of training on the inter-observer reliability of lameness scoring in dairy cattle', *Animal Welfare*, 16, pp. 131–133.

Marchant, J.N., Rudd, A.R. and Broom, D.M. (1997) 'The effects of housing on heart rate of gestating sows during specific behaviours', *Applied Animal Behaviour Science*, 55, pp. 67–78.

Martin, P. and Bateson, P. (1992) *Measuring Behaviour*. Cambridge University Press, Cambridge.

Martin, P. and Bateson, P. (1993) *Measuring Behaviour*, 2nd edn. Cambridge University Press, Cambridge.

Mattiello, S., Arduino, D., Tosi, M.V. and Carenzi, C. (2005) 'Survey on housing, management and welfare of dairy cattle in tie-stalls in western Italian Alps', *Acta Agriculturae Scandinavica Section A: Animal Science*, 55(1), pp. 31–39.

McKenna, D.R., Roebert, D.L., Bates, P.K., Schmidt, T.B., Hale, D.S., Griffin, D.B., Savell, J.W., Brooks, J.C., Morgan, J.B., Montgomery, T.H., Belk, K.E. and Smith, G.C. (2002) 'National beef quality audit-2000: survey of targeted cattle and carcass characteristics related to quality, quantity, and value of fed steers and heifers', *Journal of Animal Science*, 80, pp. 1212–1222.

McNally, P.W. and Warriss, P.D. (1996) 'Recent bruising in cattle at abattoirs', *Veterinary Record*, 138, pp. 126–128.

McNally, P.W. and Warriss, P.D. (1997) 'Prevalcence of carcase bruising and stick-marking in cattle bought from different auction markets', *Veterinary Record*, 140, pp. 231–232.

Menke, C., Waiblinger, S., Fölsch, D.W. and Wiepkema, P.R. (1999) 'Social behaviour and injuries of horned dairy cows in loose housing systems', *Animal Welfare*, 8, pp. 243–258.

Menke, C.A. (1996) *Laufstallhaltung mit behornten Milchkühen*. Dissertation, Eidgenössische technische Hochschule, Zürich.

Miller, C.P., Wood-Gush, D.G.M. and Martin, P. (1986) 'The effect of rearing systems on the responses of calves to novelty', *Biology of Behaviour*, 11, pp. 50–60.

Minero, M., Tosi, M.V., Canali, E. and Wemelsfelder, F. (2009) 'Quantitative and qualitative assessment of the response of foals to the presence of an unfamiliar human', *Applied Animal Behaviour Science*, 116, pp. 74–81.

Mogensen, L., Nielsen, L.H., Hindhede, J., Sørensen, J.T. and Krohn, C.C. (1997) 'Effect of space allowance in deep bedding systems on resting behaviour, production, and health of dairy heifers', *Acta Agriculturae Scandinavica Section A: Animal Science*, 47, pp. 178–186.

Mounier, L., Veissier, I. and Boissy, A. (2005) 'Behavior, physiology, and performance of bulls mixed at the onset of finishing to form uniform body weight groups', *Journal of Animal Science*, 83, pp. 1696–1704.

Mounier, L., Veissier, I., Andanson, S., Delval, E. and Boissy, A. (2006) 'Mixing at the beginning of fattening moderates social buffering in beef bulls', *Applied Animal Behaviour Science*, 96, pp. 185–200.

Mowbray, L., Vittie, T. and Weary, D.M. (2003) 'Hock lesions and free-stall design: effects of stall surface', *Proceedings of the 5th International Dairy Housing Conference, Fort Worth, TX, 29–31 January*, pp. 288–295.

Mulvany, P. (1981) Dairy Cow Condition Scoring, Handout no. 4468. NIRD, Reading.

Munksgaard, L. and Simonsen, H.B. (1996) 'Behavioral and pituitary adrenal-axis responses of dairy cows to social isolation and deprivation of lying down', *Journal of Animal Science*, 74, pp. 769–778.

Munksgaard, L., Jensen, M.B., Pedersen, L.J., Hansen, S.W. and Matthews, L. (2005) 'Quantifying behavioural priorities: effects of time constraints on behaviour of dairy cows, Bos taurus', *Applied Animal Behaviour Science*, 92, pp. 3–14.

Murphy, L.B. (1978) 'The practical problems of recognizing and measuring fear and exploration behaviour in the domestic fowl', *Animal Behaviour*, 26, pp. 422–431.

Napolitano, F., Grasso, F., Bordi, A., Tripaldi, C., Saltalamacchia, F., Pacelli, C. and De Rosa, G. (2005) 'On-farm welfare assessment in dairy cattle and buffaloes: evaluation of some animal-based parameters', *Italian Journal of Animal Science*, 4, pp. 223–231.

Napolitano, F., De Rosa, G., Braghieri, A., Grasso, F., Bordi, A. and Wemelsfelder, F. (2008) 'The qualitative assessment of responsiveness to environmental challenge in horses and ponies', *Applied Animal Behaviour Science*, 109, pp. 342–356.

Newberry, R.C., Wood-Gush, D.G.M. and Hall, J.W. (1988) 'Playful behaviour of piglets', *Behavioural Processes*, 17, pp. 205–216.

Niesink, R.J.M. and van Ree, J.M. (1989) 'Involvement of opioid and dopaminergic systems in isolation-induced pinning and social grooming of young rats', *Neuropharmacology*, 28, pp. 411–418.

Noordhuizen, J.P.T.M. and Metz, J.H.M. (2005) 'Food safety, animal health and Welfare', *Livestock Production Science*, 94, pp. 51–59

Oertli, B., Jakob, P. and Friedli, K. (1994) *Erarbeitung der Grundlagen zur Prüfung von Bodenbelägen im Boxenlaufstall für Milchkühe auf Tiergerechtheit*, Interner Bericht. Bundesamt für Veterinärwesen, Tänikon.

Ofner, E., Amon, B., Amon, T. and Boxberger, J. (2000) 'Validation of the TGI 35 L 1995/96 Austrian animal needs index', in: H.J. Blokhuis and E.D. Ekkel (eds) *Improving Health and Welfare in Animal Production*, EAAP Publications 102, Wageningen Pers, Wageningen, pp. 81–87.

Ofner, E., Amon, T., Lins, M. and Amon, B. (2003) 'Correlations between the results of animal welfare assessments by the TGI 35 L Austrian animal needs index and health and behavioural parameters of cattle', *Animal Welfare*, 12(4), pp. 571–578.

Otten, W., Puppe, B., Stabenow, B., Kanitz, E., Schön, P.C., Brussow, K.P. and Nürnberg, G. (1997) 'Agonistic interactions and physiological reactions of top- and bottom-ranking pigs confronted with a familiar and an unfamiliar group: preliminary results', *Applied Animal Behaviour Science*, 55, pp. 79–90.

Page, J.K., Wulf, D.M. and Schwotzer, T.R. (2001) 'A survey of beef muscle colour and pH', *Journal of Animal Science*, 79, pp. 678–687.

Phillips, C.J. and Morris, I.D. (2002) 'The ablility of cattle to distinguish between, and their preference for, floors with different levels of friction, and their avoidance of floors contaminated with excreta', *Animal Welfare*, 11, pp. 21–29.

Pinheiro, J.C. and Bates, D.M. (2000) *Mixed-Effects Models in S and S-Plus*. Springer, New York.

Platz, S., Miller, F. and Unshelm, J. (1999) 'Auswirkungen von Haltungsmängeln auf Tiergesundheit und Wirtschaftlichkeit am Beispiel der Anbinde- und Laufstallhaltung von Milchkühen', *Berliner und Münchener Tierärztliche Wochenschrift*, 112, pp. 422–429.

Probst, J., Hillmann, E., Leiber, F. and Spengler Neff, A. (2008) 'Wie wirkt sich ein positives Handling von Mastbullen auf Verhalten und Stressreaktivität vor der Schlachtung sowie auf die Fleischqualität aus?', *KTBL-Schrift*, 471, pp. 58–68.

Radostits, O.M., Gay, C.C., Blood, D.C. and Hinchcliff, K.W. (1999) *Veterinary Medicine: A Textbook of the Diseases of Cattle, Sheep, Pigs, Goats and Horses*, 9th edn. W.B. Saunders, London.

Ramos, A. and Mormède, P. (1998) 'Stress and emotionality: a multidimensional and genetic approach', *Neuroscience and Biobehavioral Reviews*, 22, pp. 33–57.

Raussi, S., Lensink, B.J., Boissy, A., Pyykkönen, M. and Veissier, I. (2003) 'The effect of contact with conspecifics and humans on calves' behaviour and stress responses', *Animal Welfare*, 12, pp. 191–203.

Redbo, I. (1992) 'The influence of restraint on the occurrence of oral stereotypies in dairy cows', *Applied Animal Behaviour Science*, 35, pp. 115–123.

Redbo, I. and Nordblad, A. (1997) 'Stereotypies in heifers are affected by feeding regime', *Applied Animal Behaviour Science*, 53, pp. 193–202.

Redbo, I., Emanuelson, M., Lundberg, K. and Oredsson, N. (1996) 'Feeding level and oral stereotypies in dairy cows', *Animal Science*, 62, pp. 199–206.

Regula, G., Danuser, J., Spycher, B. and Wechsler, B. (2004) Health and welfare of dairy cows in different husbandry systems in Switzerland', *Preventive Veterinary Medicine*, 66(1/4), pp. 247–264

Reid, I.M., Roberts, C.J., Treacher, R.J. and Williams, L.A. (1986) 'Effect of body condition at calving on tissue mobilization, development of fatty liver and blood chemistry of dairy cows', *Animal Production*, 43, pp. 7–15.

Reinhardt, V. (1980) 'Untersuchung zum Sozialverhalten des Rindes. Eine zweijährige Beobachtung an einer halb-wilden Rinderherde (Bos indicus)', *Tierhaltung*, 10.

Reinhardt, V. and Reinhardt, A. (1982) 'Mock fighting in cattle', *Behaviour*, 81, pp. 1–13.

Richards, M.W., Spitzer, J.C. and Warner, M.B. (1986) 'Effect of varying levels of postpartum nutrition and body condition at calving on subsequent reproductive performance in beef cattle', *Journal of Animal Science*, 62, pp. 300–306.

Rodenburg, J., House, H.K. and Anderson, N.G. (1994) 'Free stall base and bedding materials: Effect on cow comfort', in: R. Bucklin (ed.) *Dairy Systems for the 21st century: Proceedings of the 3rd International Dairy Housing Conference, 2–5 Feb., Orlando, FL*, ASAE, St. Joseph, MI, pp. 159–164.

Rodriguez-Lainz, A., Hird, D.W., Carpenter, T.E. and Read, D.H. (1996) 'Case-control study of papillomatous digital dermatitis in southern California dairy farms', *Preventive Veterinary Medicine*, 28, pp. 117–131.

Rosenberger, G. (1970) *Krankheiten des Rindes*. Paul Parey, Berlin.

Rousing, T. and Waiblinger, S. (2004) 'Evaluation of on-farm methods for testing the human-animal relationship in dairy herds with cubicle loose housing systems: test-retest and inter-observer reliability and consistency to familiarity of test person', *Applied Animal Behaviour Science*, 85, pp. 215–231.

Rousing, T. and Wemelsfelder, F. (2006) 'Qualitative assessment of social behaviour of dairy cows housed in loose housing systems', *Applied Animal Behaviour Science*, 101, pp. 40–53.

Rousing, T., Bonde, M. and Sorensen, J.T. (2000) 'Indicators for the assessment of animal welfare in a dairy cattle herd with a cubicle housing system', in: H.J. Blokhuis and E.D. Ekkel (eds) *Improving Health and Welfare in Animal Production*, EAAP Publications 102, Wageningen Pers, Wageningen, pp. 37–44.

Rousing, T., Bonde, M. and Sørensen, J.T. (2001) 'Aggregating welfare indicators into an operational welfare asessement system: a bottom up approach', *Acta Agriculturae Scandinavica Section A: Animal Science*, 30(Suppl.), pp. 53–57.

Rousing, T., Ibsen, B. and Sørensen, J.T. (2005) 'A note on: on-farm testing of the behavioural response of group-housed calves towards humans: test-retest and inter-observer reliability and effect of familiarity of test person', *Applied Animal Behaviour Science*, 94, pp. 237–243.

Rushen, J. and de Passillé, A.M. (1995) 'The motivation of non-nutritive sucking in calves, Bos taurus', *Animal Behaviour*, 49, pp. 1503–1510.

Rushen, J. and de Passillé, A.M. (2006) 'Effects of roughness and compressibility of flooring on cow locomotion', *Journal of Dairy Science*, 89(8), pp. 2965–2972.

Rushen, J., Pombourcq, E. and de Passillé, A.M. (2006) 'Local anaesthetic as a means to validate measures of lameness in dairy cows', in: M. Mendl et al. (ed.) *Proceedings of the 40th International Congress of the ISAE, Bristol*, ISAE, Edinburgh, p. 53.

Rutegård, Å. (2006) Managing Director, Swedish Meat Industry Association, personal communication, 10 February.

Ruunaniemi, J., Hautala, M. and Ahokas, J. (2005) 'Physical properties of synthetic bedding materials for free-stall dairy cow', *Agricultural and Food Science*, 14(2), pp. 134–142.

Sambraus, H.H. (1969) 'Das soziale Lecken des Rindes', *Zeitschrift für Tierpsychologie*, 26, pp. 805–810.

Sambraus, H.H. and Gotthardt, A. (1985) 'Prepuce sucking and tongue play in housed fattening bulls', *Deutsche Tierärztliche Wochenschrift*, 92, pp. 465–468.

Sato, S. (1984) 'Social licking pattern and its relationship to social dominance and live weight gain in weaned calves', *Applied Animal Behaviour Science*, 12, pp. 25–32.

Sato, S. and Tarumizu, K. (1993) 'Heart rates before, during and after allo-grooming in cattle', *Journal of Ethology*, 11, pp. 149–150.

Sato, S., Sako, S. and Maeda, A. (1991) 'Social licking patterns in cattle (Bos taurus): influence of environmental and social factors', *Applied Animal Behaviour Science*, 32, pp. 3–12.

Sato, S., Tarumizu, K. and Hatae, K. (1993) 'The influence of social factors on allogrooming in cows', *Applied Animal Behaviour Science*, 38, pp. 235–244.

SCAHAW (2001) *The Welfare of Cattle Kept for Beef Production*. European Commission, Brussels.

Scanga, J.A., Belk, K.E., Tatum, J.D., Grandin, T. and Smith, G.C. (1998) 'Factors contributing to the incidence of dark cutting beef', *Journal of Animal Science*, 76, pp. 2040–2047.

Schaub, J., Friedli, K. and Wechsler, B. (1999) *Weiche Liegematten für Milchvieh-Boxenlaufställe: Strohmatratzen und sechs Fabrikate von weichen Liegematten im Vergleich*, FAT-Bericht 529. FAT, Tänikon.

Schloeth, R. (1961) 'Das Sozialleben des Camargue Rindes', *Zeitschrift für Tierpsychologie*, 18, pp. 574–627.

Schmied, C. and Waiblinger, S. (2006a) 'Validity of different tests for assessing the cow-human relationship on dairy farms', Joint East and West Central Europe ISAE regional meeting, Celle, Germany, 18–20 May.

Schmied, C. and Waiblinger, S. (2006b) 'Handling practices during rearing are linked to the cow–human relationship', in: M. Mendl et al. (ed.) *Proceedings of the 40th International Congress of the ISAE, Bristol*, ISAE, Edinburgh, p. 179.

Schrader, L. (2002) 'Consistency of individual behavioural characteristics of dairy cows in their home pen', *Applied Animal Behaviour Science*, 77, pp. 255–266.

Schrader, L., Roth, H.R., Winterling, C., Brodmann, N., Langhans, W., Geyer, H. and Graf, B. (2001) 'The occurrence of tail tip alterations in fattening bulls kept under different husbandry conditions', *Animal Welfare*, 10, pp. 119–130.

Schreiner, D.A. and Ruegg, P.L. (2002) 'Effects of tail docking on milk quality and cow cleanliness', *Journal of Dairy Science*, 85, pp. 2503–2511.

Schreiner, D.A. and Ruegg, P.L. (2003) 'Relationship between udder and leg hygiene scores and subclinical mastitis', *Journal of Dairy Science*, 86, pp. 3460–3465.

Schukken, Y.H., Grommers, F.J., Van der Geer, D., Erb, H.N. and Brand, A. (1990) 'Risk factors for clinical mastitis in herds with low bulk milk somatic cell count: 1. Data and risk factors for all cases', *Journal of Dairy Science*, 73, pp. 3463–3471.

Schulze Westerath, H., Gygax, L., Mayer, C. and Wechsler, B. (2007) 'Leg lesions and cleanliness of finishing bulls kept in housing systems with different lying area surfaces', *Veterinary Journal*, 174, pp. 77–85.

Schulze Westerath, H., Laister, S., Winckler, C. and Knierim, U. (2009) 'Exploration as an indicator of good welfare in beef bulls: an attempt to develop a test for on-farm assessment', *Applied Animal Behaviour Science*, 116, pp. 126–133.

Scientific Veterinary Committee (1995) *Report on the Welfare of Calves*, VI/BII.2. Directorate General for Agriculture, Bruxelles.

Sedlbauer, M., Weary, D.A., Flower, F.C., Von Keyserlingk, M.A.G. and Sanderson, D.J. (2006) 'Lameness and pain in dairy cows', in: M. Mendl et al. (ed.) *Proceedings of the 40th International Congress of the ISAE, Bristol*, ISAE, Edinburgh, p. 52.

Singh, S.S., Ward, W.R., Lautenbach, K. and Murray, R.D. (1993) 'Behaviour of lame and normal dairy cows in cubicles and a straw yard', *Veterinary Record*, 133, pp. 204–208.

Singh, S.S., Ward, W.R., Hughes, J.W., Lautenbach, K. and Murray, R.D. (1994) 'Behaviour of dairy cows in a straw yard in relation to lameness', *Veterinary Record*, 135, pp. 251–253.

Sogstad, A.M., Fjeldaas, T., Osteras, O. and Plym Forshell, K. (2005) 'Prevalence of claw lesions in Norwegian dairy cattle housed in tie stalls and free stalls', *Preventative Veterinary Medicine*, 70, pp. 191–209.

Sørensen, J.T., Sandøe, P. and Halberg, H. (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*, 30(Suppl.), pp. 11–17.

Spitzer, J. (1986) 'Influences of nutrition on reproduction in beef cattle', in: D.A. Morrow (ed.) *Current Therapy in Theriogenology*, 2nd edn, W.B. Saunders, Philadelphia, PA.

Spoolder, H., De Rosa, G., Hörning, B., Waiblinger, S. and Wemelsfelder, F. (2002) 'Integrating parameters to assess on-farm welfare', *Abstracts of Presentations and Posters, 2nd International Workshop on Assessment of Animal Welfare at Farm and Group Level, University of Bristol, 4–6 September*, p. 19.

Sprecher, D.J., Hostetler, D.E. and Kaneene, J. (1997) 'A lameness scoring system that uses posture and gait to predict dairy cattle reproductive performance,' *Theriogenology*, 47, pp. 1179–1187.

Steimer, T., La Fluer, S. and Schulz, P.E. (1997) 'Neuroendocrine correlates of emotional reactivity and coping in male rats from the roman high (RHA/Verh)- and low (RLA/Verh)-avoidance lines', *Behavior Genetics*, 27, pp. 503–512.

Tapki, I. and Sahin, A. (2006) 'Comparison of the thermoregulatory behaviours of low and high producing dairy cows in a hot environment', *Applied Animal Behaviour Science*, 99(1/2), pp. 1–11.

Ter Wee, E., Wierenga, H.K., Smits, A.C. and Smits, M.C.J. (1989) *Claw and Leg Disorders in Cattle in Relation to the Design and Construction of Floors*, IVO-Report, I.V.O, Zeist.

Terosky, T.L., Wilson, L.L., Stull, C.L. and Stricklin, W.R. (1997) ‘Effects of individual housing design and size on special-fed Holstein veal calf growth performance, hematology, and carcass characteristics.’, *Journal of Animal Science*, 75, pp. 1697–1703.

Tilbrook, A.J., Hemsworth, P.H., Barnett, J.L. and Skinner, A. (1989) ‘An investigation of the social behaviour and response to humans of young cattle’, *Applied Animal Behaviour Science*, 23, pp. 107–116.

Tosi, M.V., Canali, E., Gregoretti, L., Ferrante, V., Rusconi, V.M. and Carenzi, C. (2001) ‘A descriptive analysis on welfare indicators measured in italian dairy farms: preliminary results’, *Acta Agriculturae Scandinavica Section A: Animal Science*, 30(Suppl.), pp. 69–72.

Troxler, J. (1994) *Untersuchungen zum Einfluss perforierter Böden auf Verhalten und Gesundheit von Mastkälbern. Zusammenfassung der Dissertation von Trachsel-Stettler (1989) und Diplomarbeit von Stumpf (1994)*, Vienna.

Tucker, C.B., Weary, D.M., Fraser, D. (2003) ‘Effects of three types of free-stall surfaces on preferences and stall usage by dairy cows’, *Journal of Dairy Science*, 86(2), pp. 521–529.

Tucker, C.B., Weary, D.M., Fraser, D. (2005) ‘Influence of neck-rail placement on free-stall preference, use, and cleanliness’, *Journal of Dairy Science*, 88(8), pp. 2730–2737.

Tucker, C.B., Weary, D.M., de Passillé, A.M., Campbell, B. and Rushen, J. (2006) ‘Flooring in front of the feed bunk affects feeding behavior and use of freestalls by dairy cows’, *Journal of Dairy Science*, 89(6), pp. 2065–2071.

Tucker, C.B., Zdanowicz, G. and Weary, D.M. (2006) ‘Brisket boards reduce freestall use’, *Journal of Dairy Science*, 89(7), pp. 2603–2607.

Tuyttens, F.A.M. (2005) ‘The importance of straw for pig and cattle welfare: a review’, *Applied Animal Behaviour Science*, 92(3), pp. 261–282.

Van der Tol, P.P.J., Metz, J.H.M., Noordhuizen-Stassen, E.N., Back, W., Braam, C.R. and Weijs, W.A. (2005) ‘Frictional forces required for unrestrained locomotion in dairy cattle’, *Journal of Dairy Science*, 88, pp. 615–624.

van Reenen, C.G., Engel, B., Ruis-Heutinck, L.F.M., van der Werf, J.T.N., Buist, W.G., Jones, R.B. and Blokhuis, H.J. (2004) ‘Behavioural reactivity of heifer calves in potentially alarming test situations: a multivariate and correlational analysis’, *Applied Animal Behaviour Science*, 85, pp. 11–30.

van Reenen, C.G., Hopster, H., van der Werf, J.T.N., Engel, B., Buist, W.G., Jones, R.B., Blokhuis, H.J. and Korte, S.M. (2009) ‘The benzodiazepine brotizolam reduces fear in calves exposed to a novel object test’, *Physiology and Behavior*, 96, pp. 307–314.

van Reenen, C.G., O’Connell, N.E., van der Werf, J.T.N., Korte, S.M., Hopster, H., Jones, R.B. and Blokhuis, H.J. (2005) ‘Response of calves to acute stress: Individual consistency and relations between behavioural and physiological measures’, *Physiology and Behavior*, 85, pp. 557–570.

van Vuuren, A.M., Stockhofe, N., Heeres-van der Tol, J.J., Heutinck, L.F.M. and van Reenen, C.G. (2004) ‘Effect of roughages added to the milk replacer diet of veal calves on behavior and gastric development’, in: N.P. Joshi and T.H. Herdt (eds) *Production Diseases in Farm Animals: 12th International Conference on Production*

Diseases in Farm Animals, East Lansing, MI, Wageningen Academic Publishers, Wageningen.

Veissier, I. (1993) 'Observational learning in cattle', *Applied Animal Behaviour Science*, 35, pp. 235–243.

Veissier, I., Capdeville, J. and Delval, E. (2004) 'Cubicle housing systems for cattle: comfort of dairy cows depends on cubicle adjustment', *Journal of Animal Science*, 82(11), pp. 3321–3337.

Vokey, F.J., Guard, C.L., Erb, H.N. and Galton, D.M. (2001) 'Effects of alley and stall surfaces on indices of claw and leg health in dairy cattle housed in a free-stall barn', *Journal of Dairy Science*, 84, pp. 2686–2699.

Wagner, J.J, Lusby, K.S., Oltjen, J.W., Rakestraw, J., Wettemann, R.P. and Walters, L.E. (1988) 'Carcass composition in mature Hereford cows: estimation and effect on daily metabolizable energy intake during winter', *Journal of Animal Science*, 66, pp. 603–612.

Waiblinger, S. and Menke, C. (2003) 'Influence of sample size and experimenter on reliability of measures of avoidance distance in dairy cows', *Animal Welfare*, 12, pp. 585–589.

Waiblinger, S., Knierim, U. and Winckler, C. (2001) 'The development of an epidemiologically based on-farm welfare assessment system for use with dairy cows', *Acta Agriculturae Scandinavica Section A: Animal Science*, 30(Suppl.), pp. 73–77.

Waiblinger, S., Fresdorf, A. and Spitzer, G. (2002) 'The role of social licking in cattle for conflict resolution', *Proceedings of the 1st European Conference of Behavioural Biology, Münster, 1–4 August*, p. 122.

Waiblinger, S., Menke, C. and Coleman, G. (2002) 'The relationship between attitudes, personal characteristics and behaviour of stockpeople and subsequent behaviour and production of dairy cows', *Applied Animal Behaviour Science*, 79, pp. 195–219.

Waiblinger, S., Menke, C. and Fölsch, D.W. (2003) 'Influences on the avoidance and approach behaviour of dairy cows towards humans on 35 farms', *Applied Animal Behaviour Science*, 84, pp. 23–39.

Waiblinger, S., Boivin, X., Pedersen, V., Tosi, M., Janczak, A.M., Visser, E.K. and Jones, R.B. (2006) 'Assessing the human-animal relationship in farmed species: a critical review', *Applied Animal Behaviour Science*, 101, pp. 185–242.

Waiblinger, S., Schmied, C., Minero, M., Baroli, D., Zucca, D. and Canali, E. (2007a) *Variation in Austrian and Italian Dairy Farmers' Attitudes and Handling Practices and Consequences for Handling Stress*, Report for the EEC-project Welfare Quality®, Vienna.

Waiblinger, S., Mülleder, C., Schmied, C. and Dembele, I. (2007b) 'Assessing the animals' relationship to humans in tied dairy cows: between-experimenter repeatability of measuring avoidance reactions', *Animal Welfare*, 16, pp. 143–146.

Wang, P.Y. (2004) *Investigating the Validity of Qualitative Assessments of Behaviour Using a Free Choice Profiling Approach in Chickens*. MSc Dissertation, University of Edinburgh.

Ward, W.R. (2003) 'Body condition scoring: technique and application', *Cattle Practice*, 11(3), pp. 111–115.

Ward, W.R., Hughes, J.W., Faull, W.B., Cripps, P.J., Sutherland J.P. and Sutherst, J.E. (2002) 'Observational study of temperature, moisture, pH and bacteria in straw bedding, and faecal consistency, cleanliness and mastitis in cows in four dairy herds', *Veterinary Record*, 151, pp. 199–206.

Weary, D.M. and Taszkun, I. (2000) 'Hock lesions and free-stall design', *Journal of Dairy Science*, 83, pp. 697–702.

Webb, N.G. and Nilsson, C. (1983) 'Flooring and injury: an overview', in: S.H. Baxter, M.R. Baxter and J.A.D. MacCormack (eds) *Farm Animal Housing and Welfare*, Martinus Nijhoff Publishers, Boston, pp. 226–261.

Wechsler, B., Schaub, J., Friedli, K. and Hauser, R. (2000) 'Behaviour and leg injuries in dairy cows kept in cubicle systems with straw bedding or soft lying mats', *Applied Animal Behaviour Science*, 69, pp. 189–197.

Weeks, C.A., McNally, P.W. and Warriss, P.D. (2002) 'Influence of the design of facilities at auction markets and animal handling procedures on bruising in cattle', *Veterinary Record*, 150, pp. 743–748.

Weiss, A., King, J.E. and Perkins, L. (2006) 'Personality and subjective well-being in orangutans (Pongo pygmaeus and Pongo abelii)', *Journal of Personality and Social Psychology*, 90(3), pp. 501–511.

Welp, T., Rushen, J., Kramer, D.L., Festa-Bianchet, M. and de Passillé, A.M. (2004) 'Vigilance as a measure of fear in dairy cattle', *Applied Animal Behaviour Science*, 87, pp. 1–13.

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

Wemelsfelder, F. and Birke, L. (1997) 'Environmental challenge', in: M.C. Appleby and B.O. Hughes (eds) *Animal Welfare*, CAB International, Wallingford, pp. 35–47.

Wemelsfelder, F. and Farish, M. (2004) 'Qualitative categories for the interpretation of sheep welfare: a review', *Animal Welfare*, 13, pp. 261–268.

Wemelsfelder, F., Hunter, E.A., Mendl, M.T. and Lawrence, A.B. (2000) 'The spontaneous qualitative assessment of behavioural expressions in pigs: first explorations of a novel methodology for integrative animal welfare measurement', *Applied Animal Behaviour Science*, 67, pp. 193–215.

Wemelsfelder, F., Hunter, E.A., Mendl, M.T. and Lawrence, A.B. (2001) 'Assessing the 'whole animal': a Free-Choice-Profiling approach', *Animal Behaviour*, 62, pp. 209–220.

Wemelsfelder, F., Nevison, I. and Lawrence, A.B. (2009) 'The effect of perceived environmental background on qualitative assessments of pig behaviour', *Animal Behaviour*, 78, pp. 477–484.

Whay, H.R. (2002) 'Locomotion scoring and lameness detection in dairy cattle,' *In Practice*, 24, pp. 444–449.

Whay, H.R., Waterman, A.E. and Webster, A.J.F. (1997) 'Associations between locomotion, claw lesions and nociceptive threshold in dairy heifers during the peri-partum period', *Veterinary Journal*, 154, pp. 155–161.

Whay, H.R., Main, D.C.J., Green, L.E. and Webster, A.J.F. (2003) 'Assessment of the welfare of dairy cattle using animal-based measurements: direct observations and investigation of farm records', *Veterinary Record*, 153(7), pp. 197–202.

Whay, H.R., Main, D.C.J., Green, L.E. and Webster, A.J.F. (2003a) 'Animal based measures of welfare state of dairy cattle, pigs and laying hens: consensus of expert opinion', *Animal Welfare*, 21, pp. 205–217.

Whay, H.R., Main, D.C.J., Green, L.E. and Webster, A.J.F. (2003b) 'Assessment of the welfare of dairy cattle using animal-based measurements: direct observations and investigation of farm records', *Veterinary Record*, 153, pp. 197–202.

Whitman, R.W. (1975) *Weight Change, Body Condition and Beef-cow Reproduction*. PhD Dissertation, Colorado State University, Fort Collins, CO.

Whytes, J.R., Kaus, R.K. and Newman, G.A. (1985) 'Bruising in beef cattle slaughtered at an abattoir in southern Queensland', *Australian Journal of Experimental Agriculture*, 25, pp. 727–733.

Wiepkema, P.R. (1987) 'Behavioural aspects of stress', in: P.R. Wiepkema and P.W.M. Van Adrichem (eds) *Biology of Stress in Farm Animals: An Integrative Approach*, Martinus Nijhoff Publishers, Dordrecht, pp. 113–133.

Wierenga, H.K. (1984) 'The social behaviour of dairy cows: some differences between pasture and cubicle system', in: J. Unshelm, G. van Putten, K. Zeeb (eds) *Proceedings of the International Congress on Applied Ethology in Farm Animals, Kiel*, KTBL, Darmstadt, pp. 135–138.

Wierenga, H.K. and Hopster, H. (1982) 'Die Reaktion von Milchvieh auf die Einschränkung von Liegeplätzen im Laufstall', *KTBL-Schrift*, 281, pp. 46–60.

Wildman, E.E., Jones, G.M., Wagner, P.E., Boman, R. L., Troutt, H.F. and Lesch, T.N. (1982) 'A dairy cow body condition scoring system and its relationship to selected production characteristics', *Journal of Dairy Science*, 65, p. 495.

Wilson, D.S. (1998) 'Adaptive individual differences within single populations', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 353, pp. 199–205.

Winckler, C. and Brill, G. (2004) 'Lameness prevalence and behavioural traits in cubicle housed dairy herds: a field study', *Proceedings of the 13th International Symposium and Conference on Lameness in Ruminants, Maribor*, pp. 160–161.

Winckler, C. and Willen, S. (2001) 'The reliability and repeatability of a lameness scoring system for use as an indicator of welfare in dairy cattle', *Acta Agricultura Scandinavica A*, 30(Suppl.), pp. 103–107.

Winckler, C., Bühnemann, A. and Seidel, K. (2002) 'Social behaviour of commercial dairy herds as a parameter for on-farm welfare assessment', P. Koene et al. (eds) *Proceedings of the 36th International Congress of ISAE, Wageningen*, p. 86.

Winckler, C., Capdeville, J., Gebresenbet, G., Hörning, B., Roiha, U., Tosi, M. and Waiblinger, S. (2003) 'Selection of parameters for on-farm welfare assessment protocols in cattle and buffalo', *Animal Welfare*, 12, pp. 619–624.

Winckler, C., Brinkman, J. and Glatz, J. (2007) 'Long-term consistency of selected animal-related welfare Parameters in dairy farms', *Animal Welfare*, 16, pp. 197–199.

Windschnurer, I., Schmied, C., Boivin, X. and Waiblinger, S. (2008) 'Reliability and inter-test relationship of tests for on-farm assessment of dairy cows' relationship to humans', *Applied Animal Behaviour Science*, 114, pp. 37–53.

Windschnurer, I., Boivin, X. and Waiblinger, S. (2009) 'Reliability of an avoidance distance test for the assessment of animals' responsiveness to humans and a preliminary investigation of its association with farmers' attitudes on bull fattening farms', *Applied Animal Behaviour Science*, 117, pp. 117–127.

Wood-Gush, D.G.M. and Vestergaard, K. (1989) 'Exploratory behavior and the welfare of intensively kept animals', *Journal of Agriculture Ethics*, 2, pp. 161–169.

Wood, M.T. (1977) 'Social grooming patterns in two herds of monocygotic twin dairy cows', *Animal Behaviour*, 25, pp. 635–642.

Wulf, D.M., Emnett, R.S., Leheska, J.M. and Moeller, S.J. (2002) 'Relationships among glycolytic potential, dark cutting (dark, firm, and dry) beef, and cooked beef palatability', *Journal of Animal Science*, 80, pp. 1895–1903.

Zahner, M., Schrader, L., Hauser, R., Keck, M., Langhans, W. and Wechsler, B. (2004) 'The influence of climatic conditions on physiological and behavioural parameters in dairy cows kept in open stables', *Animal Science*, 78(1), pp. 139–147.

Zerzawy, B. (1989) *Haltungsbedingte, adspektorisch und palpatorisch erfassbare Krankheiten und Abgangsursachen von Milchkühen in Abhängigkeit von den Stallverhältnissen im Liegeboxenlaufstall*. Dissertation, Justus-Liebig-Universität Giessen.

Zimmermann, A., Stauffacher, M., Langhans, W. and Würbel, H. (2001) 'Enrichment-dependent differences in novelty exploration in rats can be explained by habituation', *Behavioural Brain Research*, 121, pp. 11–20.

Zurbrigg, K., Kelton, D. and Anderson, N.M.S. (2005) 'Tie-stall design and its relationship to lameness, injury, and cleanliness on 317 Ontario dairy farms', *Journal of Dairy Science*, 88, pp. 3201–3210.

APPENDIX 1

ON-FARM ASSESSMENT OF MANAGEMENT AND HANDLING-BASED MEASURES IN CATTLE

S. Waiblinger, B. Algers, A. Butterworth and L.J. Keeling

A1.1 SUMMARY

Management-based measures refer to what the animal unit manager does on the animal unit and what management processes are used. It was foreseen that when no animal-based measure is available to check a criterion, or when such a measure is not sensitive or reliable enough, measures of resources or management will be used to check as much as possible that a given welfare criterion is met.

For cattle, 83 measures are discussed regarding their welfare relevance and a rational for inclusion or exclusion in the final list of 53 measures proposed to be considered in Welfare Quality® assessment. For those 53 measures in the final list, a short text is added proposing whether the measure can be assessed by questioning (Q), by direct observation of animals, management systems or records (A) or by a combination of both (C). For example, the maintenance of the milking machine, the amount and quality of bedding or feeding management are assessed by questioning the farmer as well as by checking records or the situation in the barn directly. Further instructions to assessors including a proposed wording of questions are provided in a preliminary assessment tool.

It is the intention that these potential management measures are used in three ways – A) As direct measures of value in the Welfare Quality® assessment. An example would be the maintenance and the cleanliness of drinking and feeding equipment. B) As measures which may be of value in the advisory aspects of the assessment in which the information produced is 'fed back' to the farmer to identify areas which may affect economic performance and the welfare of his animals. For a high acceptance of Welfare Quality® among producers we think that this advisory purpose is vital. C) As measures included in large scale on farm assessments, they offer an opportunity for epidemiological analyses aimed at furthering our knowledge of animal welfare.

Because there are a large number of potential management measures it is not assumed that all of these measures will make their way into the Welfare Quality® assessment system and so, to assist in decision making on inclusion, a ranking table is provided which is based on the assessment of the validity of the measures by the researchers involved.

A1.2 INTRODUCTION

Management-based measures refer to what the animal unit manager does on the animal unit and what management processes are used. The first and main role of management-based measures in Welfare Quality® is to address issues that can not be dealt with using animal- based measures. These measures may represent the only option for a particular dimension of animal welfare to be included in the final monitoring scheme. It is also possible that specific measures will prove to be a more feasible alternative to a single or a combination of animal based measures and so for this reason are kept in the final monitoring scheme.

A second aim of the welfare monitoring system is to be able to give feedback to the farmer/manager to help identify risks to animal welfare and causes of poor welfare so that improvement strategies can be implemented. Thus management questions may be included which we do not intend to contribute to the welfare assessment. In fact, since some of these are questions to the farmer, with little or no opportunity to check the reliability of the answer, we recommend that they do not contribute directly. Nevertheless, we recommend that these are included since they will enable Welfare Quality® to go some way towards explaining to the unit manager what factors might have contributed to the specific results from some of the animal-based measures. For example, a high level of lameness would lead an assessor to look at the response to questions related to the claw care programme of the farm and routines for cleaning the floor.

Thirdly, many of the measures are valuable for scientific reasons. Data collected in a standardised way on farm or at slaughter presents a unique opportunity for epidemiological analyses aimed at furthering our knowledge of animal welfare. It should be accepted that this is a secondary aim of the work in this deliverable and so, if there is a shortage of time when implementing the Welfare Quality® protocol, these questions are ones that should be excluded first. However, we include them since they could be a valuable scientific addition to the outputs of Welfare Quality®.

A1.3 APPROACH TAKEN

Animal-based measures are the measure of choice in Welfare Quality®, since they are independent of housing system. This is not the case for management measures which are, of course, highly dependent on the type of system. For that reason some questions are relevant in only one system, whereas others may be relevant in all systems. For this reason, in Tables A1.1 and A1.2, the system where the question can be asked is also specified. E.g. only relevant to tie-stall systems for cattle or are relevant both in tie-stall and loose housing systems We have nevertheless tried to keep to main areas of management that are common to all systems, or on decisions of the treatment of animals, which are relevant to all species and systems.

There are fewer measures that we recommend be considered for inclusion in the welfare assessment than there are measures that we suggest be used when giving advice back to the farmer/manager. This is because management is an ongoing process and although there may be paper documentation that can be examined (e.g. contract with a veterinarian), or physical evidence of a management process that can be evaluated (e.g. functioning of equipment), most measures are based on questions to the farmer/manager. It is therefore usually impossible to determine if the answer is correct. Thus we recommend that these measures are not sufficiently reliable to be used in the assessment. To maximise the reliability we have been careful in the wording of each question to avoid 'leading questions' or an implication through the formulation of the question as to what is the correct or preferred answer. In the tables, we have given the formulation of the questions, although we appreciate that these may have to be modified as well as translated into other languages. We have also given the answer categories (sometimes a scoring system) of how the answer should be recorded so that they can easily be included in the database. For this reason most answers are broken down to a 'Yes/No', or require the assessor to tick one of several options. We have not given the exact layout of the questionnaire since we presume that the order in which the questions are asked will depend very much on the decisions related to the structure of the visit itself.

The style of the questions varies and several require that the assessor evaluates the appropriateness of the answer. For this reason we have sometimes given instructions for the assessor to help them understand the exact aim of the question and give guidance on what is an appropriate answer.

Once a first list of the questions had been developed each researcher went through the list of questions and ranked them according to the following questions.

- How critical the issue addressed in the question was for animal welfare (low to high).
- The portion of life referred to in the question (short to long period of the animal's life).
- The time needed to gather the data (quick or time consuming).

- The quality/reliability of the data (poor to good).
- Whether there was a critical window during which the measurement must be taken (yes to no it could be gathered anytime).

The answers were discussed. There had already been a large amount of 'filtering' of questions and most questions ranked quite favourably (i.e. measures considered of very low value had not even been included in the initial lists). As a further confirmation of this, researchers ranked each question on 'how critical it was for welfare' and the average of their scores is also presented in the table. Presenting this average is intended to help future prioritising of questions. Table A1.2 shows a final ranking of the 53 measures for cattle which survived this process. We propose those that are ranked lowest are deleted first, even if we have already been ruthless and the list only contains questions that we firmly believe are important to include in the scheme. As a further aid in the development of the operational protocol we have listed whenever we think one of our measures relates to another measure. There are several measures which are closely related to a resource based measure in the area e.g. straw is a resource, but how often it is provided and whether it is clean or not is related to management. Some questions already relate to animal-based measures, especially those related to health.

There is some concern among those working with the animal-based measures related to fear and human animal interactions that the measures will not be independent of housing system and so may not be accepted for inclusion in the full monitoring scheme. We feel it would be a deficiency to have nothing in the monitoring system that addressed these areas. We have therefore included a series of questions related to the farmer's/manager's behaviour towards the animals and some observations of handling that we propose to be included, at least during the refinement process. If animal-based measures of fear and the human animal relationship are included in the final assessment system, then those findings can be correlated with this (perhaps more feasible) approach. If those animal-based measures are not included then these questions would at least allow investigation of the human-animal relationship. However, the more feasible, quick approaches would need some investigation with regard to reliability in case of later use in the Welfare Quality® assessment.

A1.4 MEASURES

Table A1.1 lists the potential measures to assess management and handling that we think are important for inclusion in the welfare assessment tool, together with measures not recommended for further use, to make the decision process transparent. The list includes all items that were discussed (in total 83 single items), i.e. also items that we decided not to include into the list of recommended measures. Out of these, 53 items are recommended and only these are given a number in the first column. In the second column it is specified whether the question is most relevant to the Welfare Quality® assessment (WQA), a question mark afterwards (WQA?) means that it is not recommended because the measure is (partly) not reliable, but some information is seen as very relevant for welfare and is not covered in any other measure. AT is used to denote that it can be used in some form of advisory tool and, finally, NO denotes that it is not to be included. The third column specifies the housing where the measures is applicable T= tying stall and L=loose housed or free range and the fourth column specifies the animal type for which the measure is relevant C=cows, F= fattening bulls / oxen and V= veal. Column five gives the item description and how to assess the measure i.e. whether it is a direct question (Q), whether it is assessed in the house or from paper records (A) or whether it is a combination of both (C). The sixth column gives instructions to the assessor and includes a very brief description of the general approach. More details about questions and assessment in the barn can be found in table 3. The seventh column gives a short description of the reason for including or disregarding this potential measure. The eighth column presents the welfare relevance as scored by the group on average; 1 = low; 5 =high. The ninth and final column gives the links to other sub tasks, that is to say to resource-based or animal-based measures are indicated.

Table A1.2 gives a ranking of the measures with respect to the perceived value to support decisions about inclusions of measures. The ranking of the 53 items was done according to: (1) the type of measure, i.e. WQA-measures are always ranked higher than AT. This is because it is recommended to not delete any of the WQA-measures (or WQA?) as they give directly relevant information on welfare that can not be found in other measures, (2) the average score of the group given to the item with respect to welfare relevance and (3) if the average score is the same, different ranks may be given according to the opinion of the responsible author for cattle. The same coding is used as for table 1 with the addition of the code (Epi) which denotes that this question may be useful for epidemiological analysis.

Table A1.3 presents the questions that we suggest are important for management-based measures for cattle and how answers can be recorded. The layout is only preliminary, but gives an idea of how such a questionnaire, when a final list is agreed upon, could be formulated. Detailed definitions of the categories of scoring system are given only for the questions we felt to be directly relevant for welfare assessment, i.e. WQA measures.

In the final section, approaches to assess handling are presented. These include; observations during handling, questions on the frequency of use of particular types of behaviour and a simple observation protocol.

TABLE A.1 List of all management and handling practices in cattle.

Number of item	WQA / AT	Housing	Animal type	Item description and how to assess	Instruction to assessor	Reason to include or disregard; welfare relevance; problems	Score	Links to other sub tasks
Calving management								
1	(WQA?) reliab-ility partly low => AT?	LT	C	C: What % of animals are calving in an appropriate environment? (0–100%)	Ask farmer 'What % of animals calve in the herd in the barn, on pasture, in calving box, in tie-stall'. Check location of calving for appropriateness, give total %.	Clean, dry calving area is important for calf health (avoiding navel infection) Disturbance (chasing up) during calving can cause an increase in calving time & risk of complication Isolation is a strong stressor for cattle – causes stress, consequently may negatively affect calving Thus it should go into WQA, however, % can hardly be checked, only if no appropriate environment measure exists could it be used	3,5	Resource – based. Calving box size, existence
2	AT	LT	C	Q: Does the calving management strategy seem to avoid problems? (yes/no)	Ask farmer about % of observed, % of assisted calvings,% major assistance and decide about effectiveness of strategy	Disturbance during calving can increase time & risk of complication; wrong assistance – cause injuries, calf mortality increases; BUT: difficult to interpret, toget the information exactly and difficult to 'integrate' the information, hence should be AT	4	Animal-based
	NO	LT	C	% of difficult calvings	-	Difficult calvings directly impact on welfare (pain, injuries, higher calf mortality), but definition questionable; depends on housing and management of inseminations	4,7	Animal-based
	NO	LT	C	% of stillbirths		Related to health/stress	3	Animal-based

TABLE A.1 CONT. List of all management and handling practices in cattle.

Number of item	WQA / AT	Housing	Animal type	Item description and how to assess	Instruction to assessor	Reason to include or disregard; welfare relevance; problems	Score	Links to other sub tasks
	NO	LT	C	Is there appropriate care for newborn calves (except colostrum)?		Measures calf welfare, but so far WQ does not assess calf welfare, just cow welfare	2,3	Animal-based. Infected navels
	NO		C	Is it guaranteed / cared for sufficient colostral intake?		Colostral intake is important for passive immunity and thus disease prevention, BUT calf welfare however: gives general information about health care	4,5	
Hygiene / Cleaning routine								
3	AT		C V F	C: Would you assess the cleaning strategy in running areas as appropriate? 1–5	Ask farmer for frequency and type of cleaning and check cleanliness in the barn; decide for appropriateness	Dry floors are better for claw health, minimizing claw's contact to slurry => risk factor for claw diseases, lameness	2,5	
4	AT		C V F	C: Would you assess the litter material /bedding in cubicles as well managed? 1–5	Ask farmer for cleaning frequency, littering frequency and amount and check cleanliness and litter in the barn; decide if good managed	Dirty, wet litter increases risk of infection/skin lesions, negative for claw health, reduces thermoisolating properties	3,1	
5	AT	LT F V	C V F	C: Would you assess the hygiene management of hospital pens and calving pens as sufficient?	Ask farmer about cleaning and disinfecting hospital pens and check hospital pens as well as existence of disinfecting agents; decide on appropriateness	Cleaning / desinfecting of calving pens and especially hospital pens after use is a necessary for disease prevention; clean environment important during calving	3,1	
6	AT	LT	C V F	C: Is the hygiene strategy for new bought animals appropriate?	Ask farmer about quarantine and health check; decide about appropriateness	Use of quarantine as well as health checks of new animals reduces risk of disease spreading	2,8	
7	AT	LT	C V F	C: % of animals bought in per year?	Question and check records	Higher % of bought in animals increases risk of diseases (as well as social disturbance if introduced in a group)	2,5	
8	AT	LT	V F	C: From how many different resources are animals bought?	Question and check records	Animals from different resources increases the risk of infectious diseases.	2,7	

TABLE A.1 CONT. List of all management and handling practices in cattle.

Number of item	WQA / AT	Housing	Animal type	Item description and how to assess	Instruction to assessor	Reason to include or disregard; welfare relevance; problems	Score	Links to other sub tasks
Health management strategies – routine health care - mangement of sick animals								
	NO		C V F	Existence of hospital pens?		Not included because resource based, should be checked	4,1	Resource-based
9	WQA	L	C V F	Use of hospital pens: Are obviously sick animals observed in the herd?	Check herd for animals that are obviously sick and need other conditions (e.g. littered area, easy access to feed…without social defeat; e.g. high-grade lame animals, weak animals)	Sick animals often suffer from difficult access to resources due to difficulties in walking and/or in social interactions; need for good lying area, quietness; Needs well-trained assessors	4,5	
	NO	L	C V F	Use of hospital pens: Are sick animals put in hospital pens? Criteria for putting in hospital pen. (Question to farmer)	Ask farmer for littering, cleaning frequency and check state of hospital pen in use;	Important, but not reliable; therefore WQA relies on question 9		
10	AT	LT	C V F	C: Is the hospital pen well managed?	Ask farmer and check hospital pen in use	Clean, dry, soft area important to support recovery of sick animals and avoid additional suffering	4,1	Resource-based
	NO	LT	C V F	Is the hospital pen appropriate (visual contact to conspecifics (except highly transmissible diseases); climate, size?		Important aspect but resource measure so not included. But would be possible to combine with 14	3,5	Resource-based
11	AT	LT	C V F	C: Is there evidence for an effective health strategy?	Ask farmer about their health strategy. Compare to assessment in the herd.	The health strategy/ speed of treatment and of getting the vet etc. is important for welfare. However, it is difficult to assess and should be reflected in animal-based measures. Nevertheless interesting for AT and Epi; Could be compared with animal-based health measures in the end for use in AT	4,2	Animal-based
	NO		C V F	Culling of sick animals		Waiting too long until culling is relevant for welfare. However, it is difficult to get the information needed	4,6	

TABLE A.1 CONT. List of all management and handling practices in cattle.

Number of item	WQA / AT	Housing	Animal type	Item description and how to assess	Instruction to assessor	Reason to include or disregard; welfare relevance; problems	Score	Links to other sub tasks
	NO		C V F	Transport of sick animals		Very relevant for welfare but low reliability	4,6	
12	AT	LT	C V F	C: Does an actual herd health and welfare plan exist?	Ask farmer and check the plan and its actuality.	A herd health and welfare plan can be a useful tool in health prevention	3	
Preventive measures / regular care								
13	AT	LT	C	Q: Does the claw trimming management seem appropriate?	Q: When and how often are claws trimmed?	Claw trimming is important for claw health and must be appropriate to environmental conditions	3,3	Animal-based
14	AT	LT	C	C: Claw trimming – crush available?	Ask farmer about availability of a crush, check in barn	If available on farm, the farmer may quicker claw trim when necessary	3,3	
15	AT	LT	C F V	C: Are books kept soundly? 1–5	Ask farmer about records and check	Sound records may (i) be indicators of sound management and (ii) help the farmer to realistically assess the health status and supports to take measures	2,8	
16	AT	LT	C	Q: Do you regularly control / palpate the udder of dry cows / of heifers? no/ yes	Ask farmer	Regular control of udders is important for early detection of mastitis; further, this contact may improve the cow-human relationship / reduce fear of humans	3,5	
17	AT	LT	C	Q: Do you control SCC before drying off cows? no/yes	Ask farmer	Control of SCC before drying off is important to detect subclinical mastitis and to control mastitis; other dry off strategies (slow, abrupt; antibiotics) may be important, but not clear what is best, so that those were not included	2,5	

TABLE A.1 CONT. List of all management and handling practices in cattle.

Number of item	WQA / AT	Housing	Animal type	Item description and how to assess	Instruction to assessor	Reason to include or disregard; welfare relevance; problems	Score	Links to other sub tasks
18	AT	LT	C	C: Is the milking machine maintained regularly? (Contract with company?) no/yes	Ask farmer about maintenance and ckeck contracts with company	Proper function of the milking machine is important for mastitis prevention and welfare (e.g. no leakage current).	3,6	
19	AT	LT	C	C: Indications that teat rubbers are changed regularly? no/yes			3,3	
Management Feeding								
20	WQA	LT	C V F	C: Are drinkers kept clean? no/partly/ yes	Ask and check	Clean water is important for hygienic reasons as well as cattle prefer clean water	3,1	
21	WQA	LT	C V F	C: Are feeding devices kept clean? no/ partly/yes	Ask and check	Clean feed is important for hygienic reasons as well as cattle avoid feeding dirty food	3,1	
22	WQA	LT	C V F	C: Is basic feed always freely available and contains enough roughage? no/ partly/ yes	Ask and check	Cattle have long feeding times throughout the day and need enough roughage for prevention of disease and behavioural disturbances	4,1	
23	AT	L	C V F	C: Is competition for food reduced as much as possible? no/ partly/ yes	Ask farmer about feeding practices and check in the barn	Competition during feeding causes agonistic interaction and may cause stress and impair individual intake and provision	4	Animal-based (social behaviour agonistic.; health, body condition score
	NO	L	C V F	Is competition for concentrate reduced as much as possible?		Included partly into 23; because overlaps with resource-based and animal-based (e.g. agonistic interactions at concentrate feeders)	3,1	Animal-based; resource-based (e.g. automatic feeder)
24	AT		C V F	C: How often is food analysed? number (every ? year)		For proper calculation of rations a knowledge of food composition helps	2,1	

TABLE A.1 CONT. List of all management and handling practices in cattle.

Number of item	WQA / AT	Housing	Animal type	Item description and how to assess	Instruction to assessor	Reason to include or disregard; welfare relevance; problems	Score	Links to other sub tasks
25	AT		C V F	C: How often are ratios calculated? number (every ? year)		To guarantee proper provision of nutrients, calculation of ratios is necessary	2,3	
Management with respect to comfort behaviour								
26	WQA	L	C V F	A: Are there functioning brushes available in appropriate amount? no/yes		Possibilities to perform appropriate comfort behaviour for cleaning the coat enhances the welfare of the animals	3,1	Resource-based
27	WQA, AT	LT	C V F	C: Access to pasture – how often, how many hours? days/year, hours/day	Ask amount and check plausibility by checking records, conditions of pasture & eventually release animals on pasture	Access to pasture elicits signs of positive emotion (locomotor play behaviour, social play) and has health benefits when long enough. Problem: not clear if reliability high enough to let it contribute to the WQA; however, in case of clear positive or negative indications it may (i.e. cows obviously used to pasture, pasture access paths look used)	3,5	Resource-based (existence)
28	WQA, AT	LT	C V F	C: Access to outside run – how often, how many hours? days/year, hours/day	Ask amount and check plausibility by checking records, conditions of outside run & eventually release animals in the outside run	Access to outside yard generally offers opportunity for positive emotions (sun-bathing); has beneficial health effects in T if long enough, probably also in L; decrease of agonistic interaction in L; Problem: not clear if reliability high enough to let it contribute to the WQA; however, in case of clear positive or negative indications it may (i.e. cows obviously are used to the outside run, the run looks used or not, cows exited when opened)	3,5	Resource – based (existence, size, orientation of outside run)

TABLE A.1 CONT. List of all management and handling practices in cattle.

Number of item	WQA / AT	Housing	Animal type	Item description and how to assess	Instruction to assessor	Reason to include or disregard; welfare relevance; problems	Score	Links to other sub tasks
29	WQA	T	C F	A: Is the electrical cow trainer managed acceptable? no/partly/yes		Electrical cow trainers can cause stress and behavioural changes, especially when badly managed; difficult to find indications of this by AB parameter, thus inclusion as WQA measure here	4,7	Resource-based (existence)
Management of group dynamics (social environment)								
30	AT (WQA?)	LT	C	Q: Cow-calf-contact: When are calves separated from the mother? time of contact	Ask: farmer 'When do you separate the calf from the cow on average?'; Is the cow allowed to lick and suck the calf'	Oxytocin improves release of placenta and later of uterus-involution; (also positive effects on calf immune-globuline level)	3,6	
31	AT	L	C	Q: Measures with (nervous) cows in heat? no/yes	Ask farmer:	Cows in heat in herd: risk of injuries due to jumping up (especially in cubicles) and increased social agonistic interactions	3,1	
32	AT	L	C	C: Horned cows: Are top of horns rounded? no/ yes	Check horns, especially of young cows and ask farmer if he/she rounds the horns	Sharp horns can more easily cause (severe) injuries	2,6	
33	AT		C	C: Measures with aggressive animals? no/yes	Ask farmer for measures in case of aggressive animals and check in barn if there is any	Single aggressive animals can cause stress for several others, increase the amount of agonistic interactions considerably	3,5	
34	AT	T	C F	Q: Is social stress of animals in tie stalls minimized? no/partly/yes		Subdominant cows can suffer from social stress with dominant, aggressive neighbour cows	3,7	
35	AT (WQA?)		C V F	C: Are changes in group composition rare? 1-5	Ask farmer for group changes and check in barn	Regroupings cause stress, enhances fighting (risk of injuries) due to new establishment or reestablishment of dominance relationships; stable groups enable stable social bonds with positive effects	4	

TABLE A.1 CONT. List of all management and handling practices in cattle.

Number of item	WQA / AT	Housing	Animal type	Item description and how to assess	Instruction to assessor	Reason to include or disregard; welfare relevance; problems	Score	Links to other sub tasks
	AT		C	Duration of separation of dry cows from herd	included in 40	The frequency of fights when coming back in herd increases with time of separation, starting with about 2 weeks	2,8	
36	AT		C	A: Average age of cow herd	Check records for cows' age	Older animals seem to contribute to a stable dominance hierarchy reducing social stress; further is says something about health problems; however, it is confounded by culling strategy (e.g. how important is early milk yield) and thus only interesting for AT and Epi	3,5	
37	AT	L	C	Q: When (age, stage of pregnancy) are heifers integrated into the herd?	Ask farmer		3	
38	AT	L	C	Q: How are new animals (heifers, primiparous cows, bought in cows) integrated into cow herd? Is stress minimised? no/partly/yes	Ask farmer about course of integration	Integration into a new herd (and environment) is stressful for the animals and measures to minimize this enhance welfare	3,5	
39	AT	T	C	Q: Are new animals put besides especially tolerant cows? no/yes	Ask farmer	If social stress is minimized the adaption to the new environment is eased	3,5	
40	AT	LT	C	C: How intensive is the contact between young stock and cows? 1–5	Ask farmer and check in the barn	Early integration / contact may reduce agonistic interactions during integration	3,3	
Management of reproduction / breeding & genetic strategies								
	NO	LT	C	Breeding strategy? For cows / insemination or bulls		Relevance not clear	1,6	
	NO	LT	C	What criteria do you have to keep a cow for breeding, to cull her?		Welfare relevance not clear	2,6	

TABLE A.1 CONT. List of all management and handling practices in cattle.

Number of item	WQA / AT	Housing	Animal type	Item description and how to assess	Instruction to assessor	Reason to include or disregard; welfare relevance; problems	Score	Links to other sub tasks
41	AT	LT	C	Q: Do you select for docile, non-aggressive animals? no/partly/yes	Ask farmer about selection strategies	Selection for docile, non-aggressive animals may reduce handling stress and agonistic interactions	3,5	
	NO	LT	C	When are heifers/ young stock inseminated the first time		Relevant, but better to check animal-based measures	2,8	Animal-based (number difficult calvings.)
	NO	LZ	C	Hormonal intervention (synchronising cows…)		There are some indications that hormonal interventions, especially for synchronisation or embryo transfer may cause welfare problems, but too little evidence to include as WQA or even AT	3	
Overall stockmanship								
	NO	LT	C V F	Decision for the housing they have – why How satisfied are you with your current system with respect to: - Technical functioning - How easy it will makes the management - Costs - If it is good for the animals - If you would have the money, in which area would you like to improve and why?		Why it might be relevant, showing the farmer's attitude to the housing system, it may be not very reliable, is not applicable for every situation and difficult to interpret		
Contact to animals								
	AT		C V F	Q: Number of milkers / of stockpeople		Lower number of milkers / stockpeople enhances constancy for the animals, included in 42	3	
	NO		C V F	Division of tasks: Are people constantly doing the same tasks?		No inclusion because of low / unknown welfare relevance	2	

TABLE A.1 CONT. List of all management and handling practices in cattle.

Number of item	WQA / AT	Housing	Animal type	Item description and how to assess	Instruction to assessor	Reason to include or disregard; welfare relevance; problems	Score	Links to other sub tasks
	AT		C V F	How many 'main stockpeople' are on the farm and how constant are the 'main stockpeople'?		Combined with 42	3,8	
	AT (WQA?)		C V F	Do you have 'non-regular' helpers often		Combined with 42 Cows show higher heart rate, less milk yield when milked by replacement-milkers	3,5	
	AT		C V F	Frequency of personnel change?		combined with 42	4	
42	AT	LT	C F V	Q: How consistent is personnel and care?		Consistency in caretakers enhances the predictability / controllability for the animals and enhances the chance of intensive stockperson-animal contact, problem recognition and solving	3,5	
	NO	LT	C	How often on average is each individual cow brushed? number per week/year		Brushing of cows decreases fear of humans except brushing is only performed to remove clumped dirt on the flank, especially in tied cows; further it is done rarely in L, thus no inclusion	3	
	NO	LT	C	Individual recognition of animals? How many animals (in %) does the stockperson know the name (or number)?		Too difficult to assess properly	4	
	NO	LT	C V F	How often do you walk through the barn for control of animals outside routine work in the barn?		Difficult to interpret	3,5	
Problem awareness – problem solving / Maintenance of equipment								
43	WQA	LT	C V F	A: Do drinkers function well and are in good condition? 1-5	Check in barn		4,1	

TABLE A.1 CONT. List of all management and handling practices in cattle.

Number of item	WQA / AT	Housing	Animal type	Item description and how to assess	Instruction to assessor	Reason to include or disregard; welfare relevance; problems	Score	Links to other sub tasks
44	WQA	L	C V F	A: Do concentrate feeders function well and are well managed? 1–5			3,8	Animal-based
45	WQA	LT	C V F	C: Is equipment (except drinkers, concentrate feeders) functioning well and in good condition? 1–5			4	Animal-based
46	WQA	T	C V F	Is the tying system well managed? no partly yes			3,8	
47	AT	LT	C V F	Does the farmer recognise and solve problems with housing equipment immediately? no/partly/yes		Immediate repair / change of defect housing equipment is important to avoid injuries and welfare problems (guarantee for water provision etc.)	4	
Mutilations								
48	WQA	LT	F	C: Is pain due to castration minimized / avoided? 1-5	Ask farmer and check animals and records	Castration causes pain but may be necessary in some systems (e.g. free range with females)	4,6	Animal-based. (number castrated)
		LT	C V F	How many animals are dehorned?		Dehorning/disbudding is very painful, but common especially in dairy cow loose housing, though problems avoided by appropriate stockmanship, housing, management combined with 49	3,3	
49	WQA	LT	C V F	C: Are pain and long-term effects due to dehorning minimized? 1-5	Ask farmer and check animals and records	By use of anaesthetics and analgesics, disbudding in early age (week 3) pain can be minimized; acids not to be used due to sound application more difficult; later dehorning takes long time healing (up to 3 months and higher risk of infection of the frontal sinus; also risk of neuroma with chronic pain is higher; social behaviour is changed for weeks	4,6	Animal-based (number dehorned)

TABLE A.1 CONT. List of all management and handling practices in cattle.

Number of item	WQA / AT	Housing	Animal type	Item description and how to assess	Instruction to assessor	Reason to include or disregard; welfare relevance; problems	Score	Links to other sub tasks
50	WQA	LT	C F	C: Are pain and long-term effects due to tail-docking minimized? 1-5	Ask farmer and check animals and records	Tail docking in cows done for milker's comfort (no dirty tail swinging), but is not performed in dairy cows in Europe (as far as I know), clear negative effects on welfare of animals (fly protection inhibited) shown when tail docked above hair tuft; tail docking in fattening bulls is done in intensive systems to avoid tail tip necrosis; short tail-docking (maximum 5 cm, hair tuft still exists) sufficient to prevent this problem	4,6	Animal-based– (number tail docked, how tail docked)
Loading on farm/transport								
		LT	C V F	Are there loading facilities on farm?	Aks farmer and check in house	Special loading facilities can ease loading and reduce stress – but not included because resource-based	4	Resource-based
		LT	C V F	Loading strategies		difficult to assess, situation dependent, but see 51	3,7	
51	AT	LT	C V F	Q: Are preparations taken to minimize stress during loading?		Preparations before loading (e.g. save corridor, lightening of corridor, feeding regime…) can reduce stress during loading and transport, but it is difficult to assess as situation dependent	3,5	
52	AT	LT	C V F	Q: Do you use special devices for moving animals (electric prod, stick)	Ask farmer	The use of electric prods causes stress, reliability may be low	4	
	NO	LT	C V F	Time to load (fatteners, calves)		Depends too much on situation, time alone does not say enough	3,2	

TABLE A.1 CONT. List of all management and handling practices in cattle.

Number of item	WQA / AT	Housing	Animal type	Item description and how to assess	Instruction to assessor	Reason to include or disregard; welfare relevance; problems	Score	Links to other sub tasks
53	(WQA?), AT	LT	C V F	Q: Are groups of or single unfamiliar animals mixed during transport?	Ask farmer	Mixing of unfamiliar animals causes stress and fights; probably could be observed during unloading at slaughter	4	Slaughter
	NO	LT	C V F	How long is travel to slaughterhouse?		Travel time can be important if too long but depends much on the transport environment, road etc.	3,5	

TABLE A1.2 Ranking of management and handling practices in cattle.

Rank	Score	Number of item in Table A1.1	WQA / AT / EPI	Housing	Animal type	Item description and how to assess
1	4,6	48	WQA	LT	F	C: Is pain due to castration minimized / avoided?
1	4,6	49	WQA	LT	C V F	C: Are pain and long-term effects due to dehorning minimized?
1	4,6	50	WQA	LT	C F	C: Are pain and long-term effects due to tail-docking minimized?
2	4,5	9	WQA	L	C V F	A: Use of hospital pens: Are obviously sick animals observed in the herd?
3	4,1	22	WQA	LT	C V F	C: Is basic feed always freely available and contains enough roughage?
3	4,1	43	WQA	LT	C V F	A: Do drinkers function well and are in good condition?
4	4	35	AT (WQA?)		C V F	C: Are changes in group composition rare?
4	4	45	WQA	LT	C V F	C: Is equipment (except drinkers, concentrate feeders) functioning well and in good condition?
4	4	53	(WQA?), AT	LT	C V F	Q: Are groups of or single unfamiliar animals mixed during transport?
5	3,8	44	WQA	L	C V F	A: Do concentrate feeders function well and are well managed?
5	3,8	46	WQA	T	C V F	A: Is the tying system well managed?
6	3,6	30	AT (WQA?)	LT	C	Q: Cow-calf-contact: When are calves separated from the mother?
7	3,5	1	(WQA?) AT	LT	C	C: What % of animals are calving in an appropriate environment?
7	3,5	27	WQA, AT	LT	C V F	C: Access to pasture – how often, how many hours?
7	3,5	28	WQA, AT	LT	C V F	C: Access to outside run – how often, how many hours?
8	3,1	20	WQA	LT	C V F	C: Are drinkers kept clean?
8	3,1	21	WQA	LT	C V F	C: Are feeding devices kept clean?
8	3,1	26	WQA	L	C V F	A: Are there functioning brushes available in appropriate amount? (possibly resource)
9	4,2	11	AT	LT	C V F	C: Is there evidence for an effective health strategy?
10	4,1	10	AT	LT	C V F	C: Is the hospital pen well managed?
11	4	23	AT	L	C V F	C: Is competition for food reduced as much as possible?
11	4	47	AT	LT	C V F	C: Does the farmer recognise and solve problems with housing equipment immediately?
12	4	2	AT	LT	C	Q: Does the calving management strategy seem to avoid problems?
13	4	52	AT	LT	C V F	Q: Do you use special devices for moving animals (electric prod, stick)
14	3,7	34	AT	T	C F	Q: Is social stress of animals in tie stalls minimized?
15	3,6	18	AT	LT	C	C: Is milking machine regularly maintained? (contract with company?)
16	3,5	16	AT, Epi	LT	C	Q: Do you regularly control / palpate the udder of dry cows / of heifers?
16	3,5	33	AT		C	C: Measures with aggressive animals?
16	3,5	36	AT, Epi		C	A: Average age of cow herd
17	3,5	42	AT	LT	C F V	Q: How consistent is personell and care?

TABLE A1.2 CONT. Ranking of management and handling practices in cattle.

Rank	Score	Number of item in Table A1.1	WQA / AT / EPI	Housing	Animal type	Item description and how to assess
18	3,5	38	AT, Epi	L	C	Q: How are new animals (heifers, primiparous cows, bought in cows) integrated into cow herd? Is stress minimised?
18	3,5	39	AT	T	C	Q: Are new animals put besides especially tolerant cows?
18	3,5	41	AT	LT	C	Q: Do you select for docile, non-aggressive animals?
18	3,5	51	AT	LT	C V F	Q: Are preparations taken to minimize stress during loading?
19	3,3	13	AT	LT	C	Q: Does the claw trimming management seem appropriate?
19	3,3	14	AT	LT	C	C: Claw trimming – crush available?
19	3,3	19	AT	LT	C	C: Indications that teat rubbers are changed regularly?
19	3,3	40	AT, Epi	LT	C	C: How intensive is the contact between young stock and cows?
20	3,1	4	AT		C V F	C: Would you assess the litter material /bedding in cubicles well managed?
20	3,1	31	AT	L	C	Q: Measures with (nervous) cows in heat?
21	3,1	5	AT	LT	C V F	C: Would you assess the hygiene management of hospital pens and calving pens sufficient?
22	3	37	AT, Epi	L	C	Q: When (age, stage of pregnancy) are heifers integrated into the herd?
23	3	12	AT, Epi	LT	C V F	C: Does an actual herd health and welfare plan exist?
24	2,8	6	AT, Epi	LT	C V F	C: Is the hygiene strategy for new bought animals appropriate?
24	2,8	15	AT, Epi	LT	C F V	C: Are books kept soundly ?
25	2,7	8	AT, Epi	LT	V F	C: From how many different resources are animals bought?
26	2,6	32	AT	L	C	C: Horned cows: Are top of horns rounded?
27	2,5	3	AT		C V F	C: Would you assess the cleaning strategy in running areas appropriate?
27	2,5	7	AT, Epi	LT	C V F	C: % of animals bought in per year?
27	2,5	17	AT, Epi	LT	C	Q: Do you control SCC before drying off cows?
28	2,3	25	AT		C V F	C: How often are ratios calculated?
29	2,1	24	AT		C V F	C: How often is food analysed?

TABLE A1.3 Assessment tool management and handling practices in cattle.

1	**appropriateness for calving**[*] What % of animals calve in the herd in the barn ....% **Yes No** on pasture ....% **Yes No** in calving box ....% **Yes No** in tie-stall ....% **Yes No** **=> What % of animal calve in an appropriate environment? ....%** [*] free movement possible; easy and save rising and lying down – soft, non-slippery floor, clean, dry; minimum disturbance by other cows; no isolation (at least visual contact to other cows), easy to control => no tie-stall; calving box with enough space (>$9m^2$ in single box, >6^2 in group box), litter deep enough, dry & clean; visual contact to other cows pasture with enough space / structuring to enable cow to separate from other cows.
2	What % of calvings do you observe? ....% What % of calvings do you intervene (assist)? ....% What % of calvings do you have major/heavy assisstence ....% **=> Does the calving management strategy seem to avoid problems? Yes No**[*] [*] yes: high % observation; low % assistence (only if necessary, not on a regular basis).
3	How often do you clean the floor (by hand, by scrabber, by tractor) ..../per day **Are floors clean (animals do not stand in dung in different areas (feeding area, running area inside, outside run, waiting room)? No 1 2 3 4 5 Yes** **=> Would you assess the floor cleaning strategy of the running area to be sufficient?** **No 1 2 3 4 5 Yes**[*] [*] yes: frequent cleaning (≥ 5 times/day in solid floors) and floors clean.
4	How often do you clean the cubicles? ..../per day How often do you litter the cubicles/lying area? ..../per day Amount of litter used? ....kg/cow/day **Is lying area sufficiently clean and dry? No 1 2 3 4 5 Yes** **Is amount of litter provided sufficient (= soft bedding)? No 1 2 3 4 5 Yes** **=> Would you assess the litter material /bedding in cubicles well managed?** **No 1 2 3 4 5 Yes** [*] well managed: if results in dry, clean, soft lying areas, enough litter used
5	clean (high pressure) disinfect Do you clean and disinfect hospital pens after use and how? Yes No Yes No Do you clean and disinfect calving pens after use and how? Yes No Yes No **Do empty hospital and calving pens look cleaned? Yes No** **Are disinfecting agents stored on the farm? Yes No** **=> Would you assess the hygiene management of hospital pens and calving pens sufficient?** **No 1 2 3 4 5 Yes** [*] yes: if high-pressure cleaned and disinfected; no: if both not.
6	Do you house bought in animals in special areas (quarantine pen)? Yes No Do you call the vet for health check of new bought animals? Yes No **=> Is the hygiene strategy for new bought animals appropriate? No 1 2 3 4 5 Yes** [*] yes: if both questions answered with yes.
7	**How much animals (in % of total herd) are bought in per year? ....%/year**
8	**From how many different resources are animals bought? average ..../per batch**
9	**Are obviously sick animals**[*] **observed in the herd? Yes No** [*] animals with severe disease having difficulties to get access to food a.o., e.g. high-grade lame animals, very weak animals.
10	How often do you litter, clean the hospital pen? ..../day **Is hospital pen in use and clean and dry? No 1 2 3 4 5 Yes** **=> Is the hospital pen well managed, i.e. deeply littered, clean, dry? No 1 2 3 4 5 Yes**

TABLE A1.3 CONT. Assessment tool management and handling practices in cattle.

11	How would you assess your own health strategy? Bad 1 2 3 4 5 Good Do you think there is room for improvement? No 1 2 3 4 5 Yes How quickly do you react when an animals starts to get ill? Waiting a while 1 2 3 4 5 6 Treat immediately **Are untreated sick animals found in the herd? No 1 2 3 4 5 Yes** **=> Is there evidence for effective health strategy? No 1 2 3 4 5 Yes*** * assessment from farmers question and impression of the herd; could be checked at the end, compared with animal-based health measures.
12	**=> Does an actual herd health and welfare plan exist? No Yes**
13	When are claws trimmed? Regularly Only if visibly too long horn Only if sick How often are claws trimmed? each animal .... times per year **=> Does the claw trimming management seem appropriate? No 1 2 3 4 5 Yes**
14	Do you have facilities for claw trimming on the farm? No Yes **Is there a claw trimming facility in the barn? No Yes** **=> Is a crush for claw trimming available on the farm? No Yes**
15	Are books kept soundly for records of Disease No Yes Treatment No Yes Losses incl. causes No Yes Production No Yes **=> Are records kept soundly? No 1 2 3 4 5 Yes**
16	**Do you regularly control / palpate the udder of dry cows / of heifers? No Yes**
17	**Do you control SCC before drying off cows? No Yes**
18	Is milking machine regularly maintained by a company? No Yes **Contract with company exists? No Yes** **=> Is milking machine regularly maintained? No Yes**
19	Do you change teat rubbers regularly? No Yes **Are teat rubbers in good condition? No 1 2 3 4 5 Yes** **=> Indications that teat rubbers are changed regularly? No Yes**
20	How often do you check cleanliness of drinkers? .... times / day How often do you clean drinkers thoroughly? every .... month **Are drinkers clean? No Partly Yes*** **=> Are drinkers kept clean? No Partly Yes** * No: drinkers not cleaned, dirty & water is dirty at moment of inspection; Partly: drinkers dirty (old dirt) – but water fresh and clean at moment of inspection OR only part of several drinkers clean, with clean water; Yes: drinkers clean & water clean at moment of inspection (some amount of fresh food allowed).
21	How often do you clean the trough? every .... day How often do you check cleanliness of /clean concentrate feeders? every .... day/week NA **Is feeding table clean? No Partly Yes*** **Are concentrate feeders clean? No Partly Yes**** **=> Are feeding devices kept clean? No Partly Yes***** * No: old feed, dung, dirt, stones laying on parts of feeding table; no regular cleaning of trough; Partly: no regular cleaning, but feed trough found to be clean; Yes: trough cleaned each time before feeding; no dirt and old food rests on feeding table. ** No: old feed rests, dung at all concentrate feeders; Partly: some concentrate feeders not clean; Yes: all cf clean. *** No: feeding table and cf ‘no’; Partly: one ‘no’, one ‘yes’; Yes: feeding table and cf ‘yes’.

TABLE A1.3 CONT. Assessment tool management and handling practices in cattle.

22	Do you feed ad libitum? No Yes Do you expect feed rests to stay? No Yes When do you feed the cows? at ……… (time if regular; basis for decision if irregulat) Do you push feed closer between feeding times? No Yes if yes, how often? **Is feed available and seem to be sufficient to next feeding time? No Partly Yes** (take into account feeding time given by the farmer) **Feed contains enough roughage No Yes** (hay, good silage in ration or extra available; not cut too much in case of TMR) **=> Is basic feed always freely available and contains enough roughage? No Partly Yes*** * Yes: fed ad lib with feed rests, feed pushed forward between feeding times and food available at time of control seem to be sufficient until next feeding time, enough roughage; Partly: basic feed always freely available, but roughage contain at least questionable; No: no feeding ad libitum, no or very few food availableat time of control.
23	Do you feed different basic food or just one (e.g. TMR)? Different Only one Do you distribute different food throughout all the feeding places or do you feed different feed at different places? Special food-special places Equal distribution Do you offer roughage additionally (hayrack) No Yes Are cows restrained in the feeding rack for some time? No Yes Are cows restrained in the feeding rack when concentrate or other preferred, limited food is offered? No Yes **Is the food of same quality throughout the feeding places? No Yes** Question 26 answered with yes – basic feed freely available No Yes **=> Is competition for feed reduced as much as possible? No Partly Yes*** * Answers on the right side support reduction of competition; Yes if all answers on the right side; No if all on the left; Partly if mixed.
24	Do you send food for analysis and if yes, how often? No Yes every ……… year **Check records of results of analysis – time lag to last report** **=> How often is food analysed? Never Every ……… year**
25	Do you calculate exact rations and if yes, how often? No Yes every ……… year **Check records of results of calculation of ratios – time lag to last time, frequency** **=> How often are ratios calculated? Never Every ……… year**
26	**Are there functioning brushes available in appropriate amount? No Yes** yes: brush still bristled, if electric brush – functioning
27	**How long do cows have access to pasture on average? ……… days/year (0–365)** **……… hours/day**
28	**How long do cows have access to the outside run on average? ……… days/year (0–365)** **……… hours/day**
29	**Individually adjusted distance to cows' withers of at least 5 cm? No Yes** **Timer regulating power, so that it runs only 1 times/ week? No Yes** **Cow trainer used for cows around calving? No Yes** **=> Is the electric cow trainer managed acceptable? No Partly Yes** * yes if all answers on right side, partly if one on left, two on right; no if two or three on left side.
30	Is the cow allowed to lick and suck the calf? No Yes When do you separate the calf from the cow on average? ……… min/hours/days p.p. **When are calves separated from the cow**, i.e. how long does the cow have contact with the calf (licking, suckling) on average? **……… min/hours/days**
31	**Measures with cows in heat? No** stay in herd **Yes** tether in herd in protected area bring out of herd for insemination bring to bull **NA** bull in herd
32	**Are top of horns round(ed)?** **No** **Yes** (=no sharp top also in young animals) **NA** (dehorned animals)

TABLE A1.3 CONT. Assessment tool management and handling practices in cattle.

33	What are you doing when you have an aggressive animal? nothing 'educate' not use for breeding cull or sell her NA (no aggressive animal) **Conspicious aggressive animals in the herd? Yes No** **=> Measures with aggressive animals? No Yes NA**
34	Do you change places if cows are incompatible? No Yes Do you select places of cows according to the cows relationship/interactions? No Yes How often do you change places of cows? Often 1 2 3 4 5 Never **=> Is social stress of animals in tie stalls minimized? No 1 2 3 4 5 Yes**
35	What is the % of replacement per year? ....% **check records** How many different production groups do you have? number of groups: .... **check in barn** How often do you regroup animals on average (incl. dried off cows)? .... per week/month How long stay dry cows out of herd on average? ≤ 2 week 2–4 week more **=> Are changes in group composition rare? No 1 2 3 4 5 Yes**
36	**Average age of cow herd? .... year**
37	**When (age, state of pregnancy) are heifers integrated in the herd?** ☐ Before insemination ☐ When pregnant ....weeks before calving ☐ After calving
38	More Stress Lower Stress Where do you integrate animals? In the barn On pasture Do you take special measures? (e.g. habituation to barn without herd) No Yes Do you integrate single animals or groups of animals? Which % of animals is integrated in Groups Singly **How are new animals (heifers, primiparous cows, bought in cows) integrated into cow herd?** **Is stress minimised? No Partly Yes** Yes: all answered on the right side; No: all answers on the left; Partly: answers mixed.
39	**Are new animals put besides especially tolerant cows in tie-stall systems? No Yes**
40	Where is young stock reared/housed: 1 Other farm 2 Same farm but different housing 3 Visual contact of pregnant heifers to cows 4 Physical contact of pregnant heifers to cows 5 In cow herd **Check in the barn** **=> How intensive is the contact between young stock and cows?** **No contact 1 2 3 4 5 High contact**
41	**Do you select for docile, non-aggressive animals? No Partly Yes**
42	Number of regular milkers: .... milker for .... cows Number of stockpeople working regul. with the animals .... person for .... cows/veal calves/bulls How often do caretakers change? Every .... year/month How many 'main stockpeople' are on the farm? .... for .... cows How frequent do the 'main stockpeople' change? ☐ ≥ once/year ☐ once/two years ☐ ≤ once/two years Do you have 'non-regular' helpers often No Yes **=> Is personell and care consistent? No 1 2 3 4 5 Yes** * Yes if low number of milkers/stockpeople (but enough to take care), and very low frequency of change.
43	**Water flow sufficient No Yes** **Risk for injuries at drinker No Yes** **Do drinkers function well and are in good condition? No 1 2 3 4 5 Yes*** * Yes (5) if water flow is 20l/min and no risk for injuries of all drinkers; No (1) if all drinkers have insufficient water flow or a high risk of injuries; 2, 3, 4 if only some of the drinkers have not enough water flow and/or risk of injuries.

TABLE A1.3 CONT. Assessment tool management and handling practices in cattle.

44	**Do feeders function well** **No** **Yes (no tickling of food)** **Risk for injuries at concentrate feeder** **No** **Yes** **Do concentrate feeders function well and are well managed? No 1 2 3 4 5 Yes*** * Yes (5) if concentrate is distributed in the preset amount, no loss of concentrate, and no risk for injuries of all drinkers; No (1) if all feeders do not distribute concentrate in preset amount, loose concentrate, or have a high risk of injuries; 2,3,4 if only some of feeders show problems.
45	**Is feeding rack functioning well** **No** **Yes** **NA** **Are cubicles in good condition (no defect part with risk of causing injuries)** **No** **Yes** **NA** **Is there any other equipment not functioning or with risk of injuries? (e.g. broken iron rod)** **No** **Yes** Is there any defect equipment in the barn at the moment? No Yes if yes: since when? ……… days/weeks/months **Is equipment (except drinkers, concentrate feeders, tying system) functioning well and in good condition? No 1 2 3 4 5 Yes*** * Yes (5) if all is functioning well and no risk of injuries due to the condition of equipment; No (1) if several problems are found throughout different equipment; 2,3,4 if single to some problems are found.
46	**Gives the 'chain' enough play / room for the cow to move?** **No** **Yes** **Is the 'chain' fixated to closely to the neck of the cow?** **No** **Yes** **Is the tying system well managed? No Yes** No: both answers on the left side; Yes: enough play of the chain and roper fixation.
47	Is there any defect equipment in the barn at the moment? No Yes if yes: since when? ……… days/weeks/months **Check results of number 49 to 52 regarding maintenance/functioning of equipment** **Does the farmer recognise and solve problems with housing equipment immediately?** **No Partly Yes*** * Yes, if question 49 to 52 answered with yes; Partly if farmer reports about defect equipment but does not repair immediately; No if farmer does report to have no defect, but in 49 to 52 a defect was found.
48	How many animals are castrated? ……… % **check in barn** From whom and how is castration performed? Age ……… Method: surgery other Anaesthetics No Yes Analgesics for how many days 0 1–2 3–7 **Check records** **Is pain due to castration minimized/avoided? No 1 2 3 4 5 Yes*** * Yes (5) no castration; 4: castration in early age with anaesthetics and analgesics for 3–7 days; No (1): castration without anaesthetics/analgesics.
49	How many animals are dehorned/disbudded? ………% **check in barn** From whom and how is dehorning/disbudding performed? Age disbudding ……… dehorning ……… Method: acids thermocauter surgery Anaesthetics No Yes Analgesics for how many days 0 1–2 3–7 **Check records** **Are pain and long-term effects due to dehorning/disbudding minimized/avoided?** **No 1 2 3 4 5 Yes*** * Yes (5) no dehorning/disbudding; 4: disbudding in early age with anaesthetics, thermocauter or surgery and analgesics for 3–7 days; No (1): dehorning of most animals when older than 6 months even with anaesthetics /analgesics or dehorning, disbudding without anesthesics.

TABLE A1.3 CONT. Assessment tool management and handling practices in cattle.

50	How many animals are tail-docked? ………% **check in barn** Length of docked tail? only tail tip (up to 5 cm), hair tuft still exists more than tail tip, hair tuft docked From whom and how is tail docking performed? Age ……… Method: rubber band surgery Anaesthetics No Yes Analgesics for how many days 0 1–2 3–7 **Check records** **Are pain and long-term effects due to tail-docking minimized/avoided? No 1 2 3 4 5 Yes*** * Yes (5) no tail-docking; 4: tail-docking of only tail tip, by surgery, in early age, with anaesthetics and analgesics for 3-7 days; No (1): tail-docking of only tail tip without anaesthetics /analgesics or tail docking with hair tuft docked (even if with anesthesics).
51	**Are preparations taken to minimize stress during loading? No 1 2 3 4 5 Yes** Yes: save corridor, good lightening of corridor.
52	**Do you use special devices for moving animals (electric prod, stick)? No Yes**
53	**Are groups of or single unfamiliar animals mixed during transport? No Partly Yes**

Note: Instructions for using the questionnaire: **Questions that are in blue** are intended to be assessed in the barn; Questions that should feed into the Assessment score have background gold; Questions that would be relevant for WQA, but are (partly) not reliable enough, nevertheless should be recorded in any case are marked background light yellow; Other questions for Advisory tool or epidemiological study; Mark NA if question not appropriate to housing system.

ASSESSMENT OF HANDLING – AMOUNT AND QUALITY OF HUMAN–CATTLE INTERACTIONS

The amount and especially quality of human-animal interaction influences animal welfare by influencing the amount of stress, with potential consequences on health, and risk of injuries. In case, no animal-based measure of the human-animal relationship is included, some (simple) assessment of the quality of human-animal interactions may be included. Three possibilities exist which may – for the beginning – be combined.

1. Observation during routine handling

This is very reliable in dairy cows during milking and calves during feeding (as long as no automatic milking system or automatic feeder is used). Previous research supports validity and reliability, but the problem is a restricted time window. Further, it is quite difficult in fattening bulls.

Validity: high; reliability: high; feasibility: low to medium

2. Questions on the frequency of use of special types of behaviour

This would be a small questionnaire, which all stock people having regular contact with the animals would have to fill in.

How often do you:

Talk to animals when approaching them	Never 1 2 3 4 5 6 7 Always
Talk to animals when walking through the group	Never 1 2 3 4 5 6 7 Always
Gently touch or stroke animals when close	Never 1 2 3 4 5 6 7 Always
Wave with the hand to move them	Never 1 2 3 4 5 6 7 Always
Hit animals with the hand to move them	Never 1 2 3 4 5 6 7 Always
Shout at animals to make them move	Never 1 2 3 4 5 6 7 Always
Shout at animals to make them stop walking or stop kicking	Never 1 2 3 4 5 6 7 Always
Sse a stick to move them by visual signs	Never 1 2 3 4 5 6 7 Always
Use a stick to move them by (slightly) hitting	Never 1 2 3 4 5 6 7 Always
Use a stick to make them stand up	Never 1 2 3 4 5 6 7 Always

Validity: not know, estimated: medium; reliability: medium (test-retest) to high (inter-observer), but problem with reliability if the measures should feed into WQA since at least from the second visit the farmers should know what is wanted; feasibility: high.

3. Simple observational protocols for a rough assessment

During the inspection the assessor should stand close stand close to the animals when discussing with the stockperson(s) for at least 10 min (e.g. while filling in the management questionnaire) and during that time should additionally observe the behaviour of the farmer and note if the farmer / stockperson uses

- Gentle interactions vocal (softly talking to)
  never occurred 1 __________________ 10 often occurred
- Gentle interactions tactile (gently touching, stroking animals)
  never occurred 1 __________________ 10 often occurred
- Aversive interactions (hitting, kicking, shouting)
  never occurred 1 __________________ 10 often occurred
- Intermediate interactions (pushing animals away with low force, very slight hit, dominant talking to animals to make them stop behaviours)
  never occurred 1 __________________ 10 often occurred
- The relation of gentle to aversive
  from 0 only gentle 1 _________________ 10 only aversive (with 50% of both in the middle
- Carefulness / suddenness of movements from
  sudden movement always 1_________________10 careful, quiet movement always

Validity: not know – guess to be medium; reliability: possibly medium; feasibility: high.

APPENDIX 2

RESOURCE-BASED PARAMETERS IN CATTLE

B. Algers, E. Canali, D. Baroli and R. Westin

A2.1 SUMMARY

Welfare Quality® will develop monitoring systems for animal welfare based on both animal parameters and design parameters for use from farm to slaughter. Housing structures and environmental factors can affect the welfare of the reared animals and they can be used as risk factors for the animal-based parameters. Within this approach the assessment of design parameters is also important in order to give a feed back to farmers. Welfare assessment protocols designed for the assessment of individual farms may take into account design or environmental parameters. A series of parameters have been identified and assessed by the use of 8 experts from 6 different EU countries. The proposed protocol covers 60 potential resource measures. Simple measure descriptions are given. Tables of the potential resource measures are provided, along with short methodologies for collection of data, suggested units, and a ranking of the potential 'value' of each measure.

A2.2 INTRODUCTION

Within Welfare Quality®, sub-project 2 will develop monitoring systems for animal welfare based on animal parameters and design parameters. Part of this task was to put forward a proposal for the standardisation of some resource-based parameters in cattle from farm to slaughter.

Resource-based measures may be very useful for many reasons: To better understand what animal-based measures mean for welfare; as a substitute for animal-based measures not reliable or not feasible at farm level; as risk factors for the welfare of the animals (see EFSA report on calves). Different exiting protocols deal mainly with design parameters such as ANI Index, which is a system developed for the assessment of animal welfare on organic farms in Austria, and some data can be obtained by farm records kept by producers, but there is the necessity of a direct assessment of resource-based measures which can provide more reliable information.

Design resources include housing system, floor space, bedding material, feeding and watering, lighting and microclimate control for the on-farm period, and e.g. vehicle design, unloading facilities, lairage and stunning facilities for the transport and slaughter periods. Cattle housing systems are various. The influences of the environment in which the animals live are important for the animal-based measures which are the main tools of the Welfare Quality® assessment scheme.

Our commitment was for standardisation only and for this reason we did not carry out animal trials. The steps of our work are described in the following parts.

A2.3 APPROACH TAKEN

We carried out a review of the scientific literature on design parameters used for the assessment of the welfare in dairy cows, fattening bulls and veal calves (see reference list). According to the available literature and personal experience in preparing and using monitoring systems incorporating design parameters, the authors prepared a first table of the possible measures; the validity and feasibility of each measure were listed. These tables show the items and the reason to include them into a protocol or to disregard them. Although other criteria could have been selected the ones described appeared to be a reasonable method for ranking the measures.

A second step was to discuss these parameters with a group of 8 experts from 6 different countries within EU (France, Italy, the Netherlands, Sweden, Ireland and Austria). The members of the group were chosen for their experience in this area and tables were delivered to them. Each member of the group individually assessed and scored the 'animal welfare impact' from low to high (1-5) for each of the discussed measures. The last step was to keep the most useful measures and to rank them in relation to the experts' opinions.

After the choice of measures, particular attention was paid to the practical sampling strategy, which should suit different production systems, different concentrations of animals per group and animals with little habituation to human proximity (i.e. fattening bulls). The proposal presents the chosen sampling methods for dairy cows, fattening bulls and veal calves.

A2.4 MEASURES

The first complete list of 92 potential resource-based measures is presented in Tables A2.1 and A2.2. Tables A2.3 and A2.4 show a final ranking of the in total 60 measures which survived the expert ranking process.

All the parameters can be taken using a checklist or measurements. The sampling strategy should include samples of all different housing types on the farm. Sampling should be done across the different stages of production for cows, fattening bulls and veal calves. All production groups should be sampled for lactating dairy cows. Precisely how each measurement should be taken will vary from parameter to parameter (see Tables A2.1 and A2.2); however, the majority are assessed by counting the numbers or allocation of equipment or resources and relating that to the total number of animals in the pen (fattening bulls, veal calves) or in the production group (dairy cows). For other parameters measures should be taken (i.e. cubicles dimensions). Some parameters will need special devices or equipment in order to take them (i.e. measurement of ammonia and light intensity).

The strategy used while assessing the resources varies according to the categories of animals and size of pen and groups:

- There should be no need for the inspector to enter small pens of veal calves (i.e. containing less than 10 animals). It could be possible to check two pens simultaneously (5 or less calves)
- For larger pens, the inspector will need to enter the pen and walk through slowly from one side to the other.
- For fattening bulls resources should be checked without entering the pens, but the chosen pens should offer high visibility.
- For dairy cows the inspector should enter the stables. If the cows are divided into production groups each group should be inspected. For each group, the occurrence of equipment will be counted together with the number of animals present in the group.
- Assessments done ‘at slaughter’ should always be performed without entering pens in use.

No reliability testing was carried out in this study, as the proposal was only for standardisation. Neither have the measures been validated within this project. However, it has been recognized that housing and environmental factors can affect welfare (see references such as last EFSA reports on calves and beef cattle reported in the reference section).

The measures are feasible; however, we do have concerns about:

- fattening bulls: due to the danger of entering into the pen there is the necessity to check facilities from the outside and it could be possible that some pens can not be checked clearly from the outside (no longer a random sample of the pens, a biased choice);
- some parameters, such as calving pens, will need to be checked with the farmer because they can be present but not used.

All these parameters can be assessed at any moment of the day. However, it is possible that some facilities will be used only during certain periods of the year, such as cooling systems.

Dairy cows, fattening bulls and veal calves should be assessed, although some parameters are specific to cows only (i.e. milking parlour). For dairy cows the resources must be sampled at every production stage. If the cows are divided into production groups each group should be inspected and production stage should be identified. For fattening bulls and veal calves a minimum of 4 pens per production stage (beginning, middle and end of production stage) per building is suggested. If there are different systems of buildings, all types must be sampled. Where there are identical buildings/systems for a production stage, these should be sampled representatively (e.g. 50% from each of two buildings). Where there are multiple non-identical buildings/systems for a production stage, a representative sample from each should be taken relative to the number of animals. The assessor need not enter pens with small numbers of veal calves if these are easily viewed from the passageway. Fattening bulls resources should always be assessed without entering the pen.

Table A2.1 Comparison of different resource measures on farm. It lists the potential measures important for inclusion in the welfare assessment tool, together with measures not recommended for further use, to make the decision process transparent. Only the 39 items later recommended for inclusion are given a number in the first column. In the second column, the animal type for which the measure is relevant is specified. In the third column there is a description of the item, followed by instructions to the assessor in the fourth column. The reasons for including the measure, or not, are given in the fifth column. The sixth column presents the welfare relevance as scored by the group on average; 1 = low; 5 =high, followed by the range of scores in the next column. The eighth and final column gives the links to other sub tasks, that is to say to management-based or animal-based measures. The table is sub divided according to where the measure should be taken on the farm e.g. in the resting area, in the passages etc., or according to the topic e.g. stable climate.

Table A2.2 Comparison of different resource measures at slaughter. This table has the same columns as table 1 but it is sub divided according to areas at the slaughterhouse e.g. unloading facilities, lairage etc. Only the 19 items later recommended for inclusion are given a number in the first column.

Table A2.3 Rank ordering of measures assessed on farm (1=low rank to 4=high rank). The proposed measures are ranked separately in relation to experts opinions, shown as mean score in the tables. We feel that all can be included in the final protocol as they are valid and can be recorded rapidly. In addition to the rank and the mean score given by the experts, a short description of the measure and the instructions to the assessor are given.

Table A2.4 Rank ordering of measures assessed at slaughter. This table has the same columns as Table A2.3.

TABLE A2.1 Comparison of different resource measures on farm.

Number of item	Animal type	Item description	Instruction to assessor	Reason to include or disregard; welfare relevance, problems	Mean score	Range	Links to other sub tasks
Resting area – cubicles							
1	All	Length and width of cubicles	Take measure. Study animals and see if length and width is sufficient to allow comfortable resting and if normal behavioural pattern is performed during rising/lying.	The animals should have sufficient space to be able to rest comfortably and to assume all possible lying positions. The length of the cubicle should allow the animal to perform normal behavioural pattern during rising and lying.	4.2	(3-5)	
2	All	Type of cubicle partitions	Watch animals during rising/lying. Decide if this is performed easily or if the animals perform abnormal behavioural patterns.	Should not cause injuries during lying/rising transitions. The design of cubicle partition can also interfere with normal behavioural patterns during rising/lying.	3.1	(1-5)	
3	All	Free space in front of cubicle (insufficient/sufficient for the animal to perform normal behaviour pattern for laying/rising)	Take measure of the free space in front of cubicle. Also watch animals during rising/lying. Decide if this is performed easily or if the animals perform abnormal behavioural patterns.	In order to perform a normal behavioural pattern during rising and lying cattle need a lot of free space in front of the cubicle.	4.1	(2-5)	
4	All	Number of cubicles/animals (in case of loose housing)	Ask farmer or count number of cubicles and number of animals.	Too few cubicles may lead to that low ranked animals are not able to get access to a cubicle to rest.	4.7	(4-5)	
5	All	Type of floor (concrete, rubber...)	Check in stable	A soft, solid flooring increases comfort during resting. Non-slippery flooring is essential to prevent falls and injury.	4.3	(2-5)	Linked to animal- based measures, such as hygiene, claw health and lameness.
No		Quality of floor (soft, hard, slippery…)		Difficult to assess and compare different farms.	4.5	(3-5)	

TABLE A2.1 CONT. Comparison of different resource measures on farm.

Number of item	Animal type	Item description	Instruction to assessor	Reason to include or disregard; welfare relevance, problems	Mean score	Range	Links to other sub tasks
6	All	Type of bedding material (straw, sawdust…)	Check in stable	Can increases comfort if flooring is hard and non-insulated. Can improve hygiene but also make hygiene worse if not kept dry and clean.	3.6	(2-5)	
No		Quality of bedding material (dry, wet, clean, dirty)		Difficult to assess and compare different farms.	4.4	(3-5)	Linked to animal-based, such as animal hygiene incidence of mastitis. Also linked to management.
Resting area – other than cubicles							
7	All	Space allowance (m^2/animal)	Ask farmer or measure size of resting area. Decide if appropriate in relation to number and size of animals.	Space needed depends on age/size of animal. The animals should have sufficient of space to be able to rest comfortably and to assume all possible lying positions and to perform normal behavioural patterns during lying/rising transitions. If too crowded the risk of getting traumatic injuries such as teat lesions increases. Low ranked animals can also choose to lay in passages instead.	4.4	(3-5)	
8	All	Type of floor (concrete, rubber, deep litter, fully/ partly slatted…)	Check in stable	Soft, solid flooring increases comfort during resting. Non-slippery flooring is essential to prevent falls and injury.	4.0	(2-5)	Linked to animal-based measures, such as hygiene, claw health and lameness.

TABLE A2.1 CONT. Comparison of different resource measures on farm.

Number of item	Animal type	Item description	Instruction to assessor	Reason to include or disregard; welfare relevance, problems	Mean score	Range	Links to other sub tasks
No		Quality of floor (slipperiness, hygiene …)		Difficult to assess and compare different farms.	4.6	(2-5)	
9	All	Type of bedding material (straw, sawdust…)	Check in stable and decide on appropriateness.	Can increases comfort if flooring is hard and non-insulated. Can improve hygiene but also make hygiene worse if not kept dry and clean. For young calves provision of straw is needed to provide a suitable environment regarding thermoregulation and comfort.	2.8	(1-5)	
No		Quality of bedding material (dry, wet, clean, dirty)		Difficult to assess and compare different farms.	4.6	(3-5)	Linked to animal-based measures such as animal hygiene incidence of mastitis.
Animal passages (in case of loose housing)							
10	All	Type of floor (concrete, rubber, deep litter, fully/ partly slatted…)	Check in stable	Non-slippery flooring is essential to prevent falls and injury.	3.5	(1-5)	Linked to animal-based measures such as hygiene, claw health and lameness.
No		Quality of floor (erosive, slippery, hygiene…)		Dirty, wet and/or erosive floors can increase number of claw lesions. However difficult to assess and compare with other farms.	4.4	(3-5)	
11	All	Width of passages (sufficient/insufficient for low ranked animals to pass high ranked)	Study animals moving in alleys and see if they can easily pass each other.	If too narrow low ranked animals can have difficulties with passing higher ranked animals.	3.8	(1-5)	

TABLE A2.1 CONT. Comparison of different resource measures on farm.

Number of item	Animal type	Item description	Instruction to assessor	Reason to include or disregard; welfare relevance, problems	Mean score	Range	Links to other sub tasks
No		Outline/design of passages (possible for animals to access water and feeding ground easily)		Design will influence how animals move between feeding ground, cubicles and so on. However difficult to decide if appropriate and to compare with other farms.	4.1	(3-5)	
Enrichment							
12	Dairy	Access to brushes (in case of loose housing)	Check in stable	Access to brushes in loose housing systems increases comfort.	3.4	(1-5)	
Hospital pen							
13	All	Number of pens/animals	Count number of hospital pens and decide if appropriate in relation to herd size.	Hospital pens facilitate medical care and supervision of sick animals in large herds. Offers the animals a calm environment.	3.3	(2-5)	
14	All	Length and width	Measure pen. Decide if appropriate in relation to size of the animals.	The animals should have sufficient space to be able to rest comfortably and to assume all possible lying positions.	3.3	(1-5)	
Feeding and drinking device							
No		Space allowance at feeding ground		Sufficient space allowance is important to prevent agonistic behaviours during feeding.	4.3	(3-5)	
15	All	Number of feeder spaces/animals (in case of loose housing)	Count number of feeder spaces for roughage and decide if appropriate in relation to number of animals.	Low ranked animals can have problems with getting access to roughage if too many animals must share feeder spaces.	3.7	(2-5)	
16	All	Number of concentrate feeders/animals	Count number of concentrate feeders and decide if appropriate in relation to number of animals.	Low ranked animals can have problems with getting access to concentrate feeders if too many animals has to share one feeder station.	3.4	(2-5)	

TABLE A2.1 CONT. Comparison of different resource measures on farm.

Number of item	Animal type	Item description	Instruction to assessor	Reason to include or disregard; welfare relevance, problems	Mean score	Range	Links to other sub tasks
17	All	Number of drinkers/animals	Count number of drinkers and decide if appropriate in relation to number of animals.	Low ranked animals can have problems with getting access to drinkers if too many animals has to share one drinker.	4.6	(4-5)	
Grazing and excercise							
18	All	Regular access to grazing (year round, during summer, every day, once a week)	Ask farmer about strategy for grazing.	Pasture is the natural habitat for cattle. Grazing thus increases animal welfare.	3.6	(1-5)	
19	All	Regular access to outdoor exercise area/ paddock (year round, during summer; every day, once a week…)	Ask farmer about strategy for outdoor exercise.	Exercise important to keep legs and muscles fit in order to be able to perform normal rising and lying behaviour patterns. Abnormal behaviour during rising/lying will increase risk of traumatic udder injuries.	3.5	(1-5)	
No		Dimension of outdoor exercise area or paddock.		Animals spend a short time periods in this area only.	2.7	(1-5)	
No		Type of bed/ground in the outside area or paddock		Insufficient drainage in outdoor paddocks can cause a poor hygiene. Difficult to assess if not studied in rainy weather conditions.	3.0	(2-5)	
20	All	Regular access to indoor exercise area (every day, once a week, once a month…)	Ask farmer about possibility to indoor exercise.	Exercise important to keep legs and muscles fit in order to be able to perform normal rising and lying behaviour patterns. Abnormal behaviour during rising/lying will increase risk of traumatic udder injuries.	2.3	(1-5)	
No		Dimension of exercise area		Animals spend a short time periods in this area only.	3.0	(1-5)	

TABLE A2.1 CONT. Comparison of different resource measures on farm.

Number of item	Animal type	Item description	Instruction to assessor	Reason to include or disregard; welfare relevance, problems	Mean score	Range	Links to other sub tasks
21	All	Type of floor in indoor exercise area	Check in stable	Non-slippery flooring is essential to prevent falls and injury.	3.3	(1-5)	
Stable climate							
22	All	Temperature	Make a mean value from a few measurements taken in different places in the stable, not too close to animals or air inlets/outlets.	Preferred temperature is depending on age and amount and type of bedding. Low temperature can make calves suffer if dry straw bedding isn't provided. For older animals, excessive heat can be a problem.	3.3	(1-5)	
23	All	Relative humidity	Make a mean value from a few measurements in different places in the stable, not too close to animals or air inlets/outlets.	A high relative humidity indicates that the capacity of the ventilation system is insufficient. It also enhances the survival of different infectious agents on the floor surface and in the aerosol and makes thermo regulation more difficult. The relative humidity is related to the temperature and the amount of air ventilated. Needs special device but easy to measure.	4.3	(2-5)	
No		Type of ventilation system		Type of ventilation system is not interesting from a welfare point of view since all systems will ventilate the building.	2.7	(1-4)	
No		Amount (m3) of air ventilated per hour and animal.		Related to temperature and relative humidity. Difficult to measure.	3.7	(2-5)	
No		Air velocity		Related to chill factor. Positive during hot weather conditions. Negative during cold and/or humid conditions. Difficult to measure in general for a whole stable.	3.6	(2-5)	

TABLE A2.1 CONT. Comparison of different resource measures on farm.

Number of item	Animal type	Item description	Instruction to assessor	Reason to include or disregard; welfare relevance, problems	Mean score	Range	Links to other sub tasks
No		Chill factor (air velocity / air temperature)		Is of high importance, especially for young animals. Is however difficult to measure.	3.4	(1-5)	
No		Dust		Respirable dust can act as carriers for infectious agents present in the aerosol. High dust levels can therefore increase the risk for spreading of airborne diseases. Is however difficult to measure.	3.5	(2-5)	
24	All	Light intensity	Make a mean value from a few measurements taken in different places in the stable where animals are held.	Sufficient provision of daylight is important to keep normal daily rhythm. Total darkness can make cattle restless and easily frightened by sudden movements or noise etc. Soft lighting during night has been shown to decrease traumatic teat injuries. Needs special device but easy to measure.	3.0	(1-5)	
25	All	Noise	Make a mean value from a few measurements taken in different places in the stable where animals are held.	High noise level can cause impaired hearing, stress reactions and disturbance in acoustic communication. Needs special device but easy to measure.	2.8	(1-4)	
26	All	Extra cooling systems (sprinklers…)	Check in stable	Important in hot climate. Necessity depends on general climate.	2.9	(1-4)	

TABLE A2.1 CONT. Comparison of different resource measures on farm.

Number of item	Animal type	Item description	Instruction to assessor	Reason to include or disregard; welfare relevance, problems	Mean score	Range	Links to other sub tasks
Air quality							
27	All	Ammonia	Make a mean value from a few measurements taken in different places in the stable where animals are held.	Reflects efficiency of ventilation and manure handling system. Ammonia is an irritant gas that causes inflammation of the mucous membrane in the eye and the respiratory tract. This makes the animals more susceptible to respiratory diseases. Also keratinolytic, thus predisposing for claw lesions. Needs special device but easy to measure.	3.6	(1-5)	
No		Carbon dioxide (CO_2)		Reflects if supplies of fresh air are sufficient. Not harmful to animals unless levels are very high (>7%).	2.7	(1-5)	
No		Carbon monoxide (CO)		Reflects efficiency of ventilation and manure handling system. Ammonia levels more easy to measure.	2.8	(1-5)	
No		Hydrogen sulphide		Reflects efficiency of ventilation and manure handling system. Is highly toxic. Long term exposure will affect blood coagulation. High levels cause paralysation of respiration and thereby sudden deaths. Ammonia levels more easy to measure.	2.8	(1-5)	
No		Methane		Reflects efficiency of ventilation and manure handling system. Not directly harmful to animals. Ammonia levels more easy to measure.	3.0	(1-5)	

TABLE A2.1 CONT. Comparison of different resource measures on farm.

Number of item	Animal type	Item description	Instruction to assessor	Reason to include or disregard; welfare relevance, problems	Mean score	Range	Links to other sub tasks
Milking facilities							
28	Dairy	Dimension and design of collection area	Study the flow of animals through the collecting area during milking and decide on appropriateness.	Collecting area should be dimensioned so that overcrowding does not occur and designed in a way that the group of cows will easily walk through to the milking parlour without need of electrical prodding.	3.9	(2-5)	
29	Dairy	Floor type of collection area	Check in stable	Non-slippery flooring is essential to prevent falls and injury.	3.9	(2-5)	
No		Slope of collecting area		Relevance for welfare not clear. A large slope can make animals reluctant to walk. However possible that they will learn to pass without fear since they walk through every day.	3.1	(1-5)	
30	Dairy	Design of passages from stable to collecting area/milking parlour (number of sharp edges/ corners…)	Walk through passages and decide on appropriateness.	Design of passages affects the behaviour of animals. Sharp edges can cause bruising. Curved passages and round crowding facilitate moving of animals.	3.9	(3-5)	
31	Dairy	Design of passages from milking parlour to the stable (number of sharp edges/ corners…)	Walk through passages and decide on appropriateness.	Design of passages affects the behaviour of animals. Sharp edges can cause bruising. Curved passages and round crowding facilitate moving of animals.	3.4	(3-5)	
No		Type of milking system (robot milking, milking parlour…)	Relevance for welfare unclear.	2.3	(1-4)		
No		Access to feed during milking		Increases milking yield but low importance for welfare.	1.6	(1-3)	

TABLE A2.1 CONT. Comparison of different resource measures on farm.

Number of item	Animal type	Item description	Instruction to assessor	Reason to include or disregard; welfare relevance, problems	Mean score	Range	Links to other sub tasks
Calving pen							
32	Dairy	Number of pens/animals	Count number of calving pens.	Use of calving pens offers a calm environment during parturition. Calvings are also better supervised in calving pens than in a loose housed system.	3.4	(2-5)	
33	Dairy	Length and width	Measure pens.	Possibility to easily turn and move around will facilitate parturition.	3.1	(1-5)	
Housing of calves and veal calves							
34	Calves	Housing of calves (possibility to social interaction, keeping calves of different age separated …)	Check housing of calves and decide on appropriateness.	Social interaction reduces stress and should be provided even if kept in individual pens. Mixing of calves of different age increases spread of disease and should be avoided if possible.	3.2	(1-5)	
35	Calves	Space allowance (m2/animal)	Measure pens and count number of calves/ pen.	Important to allow animals to rest comfortably. Overcrowding can increase risk of disease outbreak.	4.2	(3-5)	
No		Location of calves within the farm (indoor, outdoor, in the same stable as adult animals, close to milking parlour…)	Important for preventive health care but difficult to compare between farms.	3.0	(1-5)		
36	Calves	Type of milk feeding device used until weaning (possible/not possible for young calves to suckle when fed milk)	Check if teat buckets or other device allowing calves to suckle during feeding is used.	Young calves have a large behavioural need to suckle. Possibility to suckle during feeding reduces suckling on pen mates and physical objects. Drinking of milk from an open bucket is a risk factor for getting milk into rumen instead of the abomasum which will make the calf sick (so called ruminal drinker).	2.4	(1-5)	

TABLE A2.1 CONT. Comparison of different resource measures on farm.

Number of item	Animal type	Item description	Instruction to assessor	Reason to include or disregard; welfare relevance, problems	Mean score	Range	Links to other sub tasks
Outdoor shelter							
37	Beef cattle	Provision of artificial or natural shelter for protection from heavy wind and rain (in case of constant grazing)	Check provision of shelter, artificial or natural, and decide if appropriate in relation to general climate.	Heavy wind and rain can cause hypothermia, especially in young animals if they can't seek shelter. Both artificial and natural shelters like close forests can be used.	2.8	(1-5)	
38	Beef cattle	Dimension of shelters (possible for all animals to seek shelter when needed)	If shelter is provided, check if dimensions are appropriate in relation to number of animals.	All animals should be able to seek shelter at the same time.	3.2	(1-5)	
No		Design of shelter (possible for low ranking animals to get access)		High ranked animals can prevent low ranked animals from getting access to artificial shelters if the entrance is too narrow.	2.8	(1-5)	
39	All	Loading ramp angle	Check loading facilities, estimate loading ramp angle and decide if appropriate.	If loading ramp angle is more than 20º, animals are reluctant to walk.	4.0	(3-5)	
No		Type of transport vehicle (1 or 2 stories, possibility for tying cattle, flexible pen arrangements…)		Not studied so far within WQ.	2.6	(1-4)	
No		Type of ventilation during transport (fans, draught…)		Not studied so far within WQ.	4.0	(3-5)	
No		Bedding material used during transport		Not studied so far within WQ.	4.3	(4-5)	
No		Access to water during transport		Not studied so far within WQ.	4.0	(3-5)	
No		Sheltered unloading		Can facilitate unloading of animals but low relevance for animal welfare.	1.6	(1-3)	

TABLE A2.2 Comparison of different resource measures at slaughter.

Number of item	Item description	Intruction to assessor	Reason to include or disregard; welfare relevance, problems	Mean score	Range	Links to other sub tasks
Unloading facilities						
No	Sheltered unloading		Can facilitate unloading of animals but low relevance for animal welfare.	1.8	(1-3)	
1	Unloading ramp angle	Measure ramp angle.	The slope should not be more than 20°. If too sharp, animals tend to turn back or start to run during unloading.	3.8	(3-5)	Closely linked to behaviour during unloading.
2	Type of ramp partition	Check ramp partition and decide on appropriateness.	Should not have sharp edges that may cause injury.	2.5	(1-3)	
Driving alley into lairage						
3	Type of partition (solid, non-solid walls...)	Walk through the driving alleys and decide on appropriateness.	Cattle are easily frightened if people or other animals seen through non-solid partition walls.	3	(1-4)	Closely linked to behaviour during unloading.
4	Presence of sharp corners/edges		Sharp corners, edges or gates can cause severe bruising.	3.6	(2-5)	
5	Floor type and quality (solid, slatted, slippery…)		Non-slippery flooring is essential to prevent falls and injury. Cattle tend to panic and become agitated when they lose their footing.	3.8	(3-5)	
6	Light conditions		Shadows or bright light may frighten the animals.	3.2	(2-5)	
7	Noise conditions		Sudden noises like air hissing and shouting is stressful and often frighten the animals.	3	(2-4)	
Lairage						
No	Type of housing (single pens, group pens, tied stall …)		Relevance for animal welfare not clear.	2.6	(2-3)	
8	Space allowance	Study animals lying down and decide if space allowance is appropriate.	Space allowance may vary depending upon weather conditions, animal sizes and varying holding times. All animals should have enough space to be able to lie down and rest comfortably.	3.6	(2-5)	

TABLE A2.2 CONT. Comparison of different resource measures at slaughter.

Number of item	Item description	Intruction to assessor	Reason to include or disregard; welfare relevance, problems	Mean score	Range	Links to other sub tasks
9	Floor type	Check in lairage and decide on appropriateness.	Soft flooring increases comfort during resting. Non-slippery flooring is essential to prevent falls and injury.	3.6	(2-4)	
10	Noise conditions	Make a mean value from a few measurements taken in different places in the stable where animals are held.	Continuous high noise from fans and other equipment is stressful for cattle. Needs a special device but is easy to measure.	3.0	(2-4)	
11	Use of bedding material	Check if bedding is used in lairage	Increases hygiene and comfort if floors are hard.	3.8	(3-5)	
14	Access to water	Check in lairage that all animals have access to water	All animals should have access to water at lairage.	4.2	(3-5)	
15	Access to feed if kept in lairage over night	Ask about feeding routines and check in stable	All animals kept in lairage over night should be fed.	2.8	(1-5)	
Driving alley to stunning						
16	Type of partition (solid, non-solid walls ...)	Walk through the driving alleys and decide on appropriateness.	The design of the driving alleys influences the behaviour of cattle. Animals are easily frightened if people walking by seen through non-solid partition walls and if shadows or bright light shines into their eyes. Sharp corners can cause bruising and a slippery floor may cause injury. If the driving alley appears as a dead end animals are very reluctant to walk forward.	3.2	(2-4)	Closely linked to behaviour during driving.
17	Number of sharp corners/edges			4.0	(3-5)	
18	Floor type and quality			3.8	(3-5)	
19	Light conditions			3.4	(2-5)	
20	Noise conditions			3.0	(2-4)	

TABLE A2.2 CONT. Comparison of different resource measures at slaughter.

Number of item	Item description	Intruction to assessor	Reason to include or disregard; welfare relevance, problems	Mean score	Range	Links to other sub tasks
Design of stun crate						
No	Level of fixation (no fixation, head fixated, neck fixated, whole animal fixated…).		Fixation of the animal facilitates performance of correct stunning. Excessive pressure can however cause pain and should therefore be avoided. Fully sensible animals should be held in a comfortable, upright position. Difficult to assess and compare.	3.2	(1-5)	
21	Floor type	Check flooring in the stun crate and decide on appropriateness.	A slippery floor may cause injury. Cattle tend to panic and become agitated when they lose their footing.	3.0	(3-3)	

TABLE A2.3 Rank ordering of measures assessed on farm.

Rank	Mean score	Item description	Intruction to assessor
4	4.7	Number of cubicles/animals (in case of loose housing)	Ask farmer or count number of cubicles and number of animals.
4	4.6	Number of drinkers/animals	Count number of drinkers and decide if appropriate in relation to number of animals.
4	4.4	Space allowance of resting area (m^2/animal)	Ask farmer or measure size of resting area. Decide if appropriate in relation to number and size of animals.
4	4.3	Relative humidity	Make a mean value from a few measurements in different places in the stable, not too close to animals or air inlets/outlets.
4	4.3	Type of floor in cubicles (concrete, rubber...)	Check in stable
4	4.2	Length and width of cubicles	Take measure. Study animals and see if length and width is sufficient to allow comfortable resting and if normal behavioural pattern is performed during rising/laying.
4	4.2	Space allowance in calf pens (m^2/animal)	Measure pens and count number of calves/ pen.
4	4.1	Free space in front of cubicle (insufficient/sufficient for the animal to perform normal behaviour pattern for laying/rising)	Take measure of the free space in front of cubicle. Also watch animals during rising/lying. Decide if this is performed easily or if the animals perform abnormal behavioural patterns.
4	4.0	Loading ramp angle	Check loading facilities, Estimate loading ramp angle and decide if appropriate.
4	4.0	Type of floor in resting area (concrete, rubber, deep litter, fully/ partly slatted…)	Check in stable
3	3.9	Dimension and design of collection area	Study the flow of animals through the collecting area during milking and decide on appropriateness.
3	3.9	Floor type of collection area	Check in stable
3	3.9	Design of passages from stable to collecting area/milking parlour (number of sharp edges/ corners…)	Walk through passages and decide on appropriateness.
3	3.8	Width of passages (sufficient/insufficient for low ranked animals to pass high ranked)	Study animals moving in alleys and see if they can easily pass each other.
3	3.7	Number of feeder spaces/animals (in case of loose housing)	Count number of feeder spaces for roughage and decide if appropriate in relation to number of animals.
3	3.6	Regular access to grazing (year round, during summer, every day, once a week…)	Ask farmer about strategy for grazing.
3	3.6	Ammonia	Make a mean value from a few measurements taken in different places in the stable where animals are held.
3	3.6	Type of bedding material in cubicles (straw, sawdust…)	Check in stable
3	3.5	Type of floor in passages (concrete, rubber, deep litter, fully/ partly slatted…)	Check in stable
3	3.5	Regular access to outdoor exercise area/ paddock (year round, during summer; every day, once a week…)	Ask farmer about strategy for outdoor exercise.

TABLE A2.3 CONT. Rank ordering of measures assessed on farm.

Rank	Mean score	Item description	Intruction to assessor
3	3.4	Number of concentrate feeders/animals	Count number of concentrate feeders and decide if appropriate in relation to number of animals.
3	3.4	Access to brushes (in case of loose housing)	Check in stable
3	3.4	Design of passages from milking parlour to the stable (number of sharp edges/ corners…)	Walk through passages and decide on appropriateness.
3	3.4	Number of calving pens/animals	Count number of calving pens.
3	3.3	Number of hospital pens/animals	Count number of hospital pens and decide if appropriate in relation to herd size.
3	3.3	Length and width of hospital pens	Measure pen. Decide if appropriate in relation to size of the animals.
3	3.3	Type of floor in indoor exercise area	Check in stable
3	3.3	Temperature	Make a mean value from a few measurements taken in different places in the stable, not too close to animals or air inlets/outlets.
3	3.2	Housing of calves (possibility to social interaction, keeping calves of different age separated …)	Check housing of calves and decide on appropriateness.
3	3.2	Dimension of shelters (possible for all animals to seek shelter when needed)	If shelter is provided, check if dimensions are appropriate in relation to number of animals.
3	3.1	Type of cubicle partitions	Watch animals during rising/lying. Decide if this is performed easily or if the animals perform abnormal behavioural patterns.
3	3.1	Length and width of calving pens	Measure pens.
3	3.0	Light intensity	Make a mean value from a few measurements taken in different places in the stable where animals are held.
3	2.9	Extra cooling systems (sprinklers…)	Check in stable
2	2.8	Noise	Make a mean value from a few measurements taken in different places in the stable where animals are held.
2	2.8	Provision of artificial or natural shelter for protection from heavy wind and rain (in case of constant grazing)	Check provision of shelter, artificial or natural, and decide if appropriate in relation to general climate.
2	2.8	Type of bedding material (straw, sawdust…)	Check in stable and decide on appropriateness.
2	2.4	Type of milk feeding device used until weaning (possible/not possible for young calves to suckle when fed milk)	Check if teat buckets or other device allowing calves to suckle during feeding is used.
2	2.3	Regular access to indoor exercise area (every day, once a week, once a month…)	Ask farmer about possibility to indoor exercise.

TABLE A2.4 Rank ordering of measures assessed at slaughter.

Rank	Mean score	Item description	Intruction to assessor
4	4.2	Access to water in lairage	Check in lairage that all animals have access to water
4	4.0	Number of sharp corners/edges in driving alleys to stun crate	Walk through the driving alleys and decide on appropriateness.
3	3.8	Floor type in driving alleys to stun crate	Walk through the driving alleys and decide on appropriateness.
3	3.8	Use of bedding material in lairage	Check if bedding is used in lairage
3	3.8	Unloading ramp angle	Measure ramp angle.
3	3.8	Floor type in driving alleys into lairage (solid, slatted, slippery…)	Walk through the driving alleys and decide on appropriateness.
3	3.6	Presence of sharp corners/edges in driving alleys into lairage.	Walk through the driving alleys and decide on appropriateness.
3	3.6	Space allowance in lairage	Study animals laying down and decide if space allowance is appropriate.
3	3.6	Floor type in lairage	Check in lairage and decide on appropriateness.
3	3.4	Light conditions in driving alleys to stun crate	Walk through the driving alleys and decide on appropriateness.
3	3.2	Light conditions in driving alleys into lairage	Walk through the driving alleys and decide on appropriateness.
3	3.2	Type of partition in driving alleys to stun crate (solid, non-solid walls ...)	Walk thorugh the driving alleys and decide on appropriateness.
3	3.0	Noise conditions in driving alleys to stun crate	Walk through the driving alleys and decide on appropriateness.
3	3.0	Floor type in stun crate	Check flooring in the stun crate and decide on appropriateness.
3	3.0	Type of partition (solid, non-solid walls...)	Walk through the driving alleys and decide on appropriateness.
3	3.0	Noise conditions in driving alleys into lairage	Walk through the driving alleys and decide on appropriateness.
3	3.0	Noise conditions in lairage	Make a mean value from a few measurements taken in different places in the stable where animals are held.
2	2.8	Access to feed if kept in lairage over night	Ask about feeding routines and check in stable
2	2.5	Type of partition at unloading ramp	Check ramp partition and decide on appropriateness.

APPENDIX 3

NOTE ON CORRELATIONS: SPEARMAN AND PEARSON CORRELATION

B. Engel and C.S. van Reenen

In a previous note, we calculated how many pairs of observations one needs for inference about correlations. To be more specific, imagine that the pairs consist of duplicate observations of the same observer on a number (n) of animals. In that case the correlation is a (scale free) measure of intra observer repeatability. Alternatively, one might imagine observations on the same animals by two different observers. In that case the correlation is a measure of reproducibility. We assume that the animals are a representative (as if random) sample. In another setting the units may be pens or cages, rather than animals.

In order to find (with a probability of 95%) that the sample correlation is significantly different from 0, while the true correlation is 0.65, we need $n = 20$ pairs of observations. In order to get a 95% confidence interval for the correlation with a width less than 0.2, while the true value of the correlation is 0.8, we need $n = 55$ pairs of observations.

These results were derived for Pearson's correlation coefficient, assuming bivariate normality for the pairs of data. The main reasons for assuming bivariate normality are:

- Under bivariate normality, there is a true correlation that we want to estimate. This true correlation is one of the parameters of the bivariate normal distribution, i.e. the true correlation is a population parameter.
- Under bivariate normality, the (true) correlation has a clear interpretation as a measure of linear dependence. Testing for a non-zero correlation is equivalent to testing for a non-zero slope in one of the associated regressions of one observation in a pair on the other observation of the same pair.

The calculation of Spearman's correlation (or Spearman's rho) is similar to the calculation of Pearson's correlation: the same calculations are made after observations are replaced by rank numbers. Spearma's rho, in it's original form, is not an estimate of a population parameter. It is possible to introduce a population parameter as if that parameter was estimated by Spearman's rho, but this is somewhat contrived and it is not quite clear what kind of dependence we are measuring. So, the idea of the width of a confidence interval less than 0.2 that leads to $n = 55$ for a Pearson correlation has no obvious extension towards Spearman's rho. The same holds for Kendall's coefficient of concordance, since that coefficient can be derived from an average of Spearman correlations.

Often Spearman's rho is used to test for independence. Assuming independence, derivation of the distribution of Spearman's rho is relatively straightforward. For dependent data however, the distribution depends on the particular dependence structure that is assumed. Although, for continuous observations, the data in the pairs of observations can be transformed towards normality, a transformation towards bivariate normality does not necessarily exist.

The relative efficiency of Spearman's rho as a test statistic for dependence compared with Pearsons'correlation, assuming bivariate normality, is high. So, it is not unreasonable to use the result of $n = 20$ for Spearman's rho as well. Actually, for $n = 20$, assuming bivariate normality, the power of the test based on Pearson's correlation is 95%, while the test based on Spearman's rho has power 93%.

There is no obvious extension of the idea of the expected width of a confidence interval for Spearman's rho. What can be done, to offer some impression of the relationship between the two coefficients, is to assume bivariate normality and study the distribution of Spearman's rho. We assume that the true correlation (in the bivariate normal distribution) is 0.8. So, the true value for Pearson's correlation is 0.8. For various values of n = the number of pairs, we simulated data (a 1000 data sets to be precise) and calculated the Pearson and Spearman correlations.

For $n = 55$ the average width of the confidence interval for Pearson's correlation is 0.2. The range of values for Spearman's rho, as measured by 2.5 and 97.5 percentile points, is 0.64 – 0.87. Let us assume that we aim for a range (between the 2.5 and 97.5 percentile points) less than 0.2. For $n = 75$, the range for Spearman's rho is 0.66 – 0.86. The corresponding range for Pearson's correlation is 0.69 – 0.87. When we increase the number of pairs to n = 150 we find a range of 0.70 – 0.85 for Spearman's rho and 0.74 – 0.85 for Pearson's correlation.

So, results for the two correlations are similar, but the left hand tail of the distribution of Spearman's rho is somewhat longer. In order to have a range of width 0.2 for Spearman's rho we need to increase the number of pairs to $n = 75$. The calculations were performed under restricted conditions; assuming bivariate normality with a true correlation of 0.8.

CONCLUSION

- Both for Pearson and Spearman correlation we advise a minimum of $n = 20$ pairs of observations. This advice was based on a power of 93 to 95% under assumption of bivariate normality and a true correlation of 0.65 for the test of independence. This of course is the poor man's choice when we are talking about repeatability or reproducibility.
- In order for the range of Spearman's rho to be smaller than 0.2, we need at least $n = 75$ pairs of observations. Calculations were performed assuming bivariate normality and a true correlation of 0.8.
- For $n = 75$ pairs of observations, with a true correlation of 0.8, the outcome of Spearman's rho may be as low as 0.66. For Pearson's correlation the outcome may be as low as 0.69. In order to have the lower end of the range close to 0.7, $n = 150$ pairs are needed for Spearman's rho.
- Here is more of a recommendation than a conclusion. In practice one might use the Spearman correlation when a normal approximation to the data does not apply. This might be the case when the data are quite discrete. The most extreme example being 0-1 data. In that case often more direct measures for repeatability or reproducibility (or association in general) can be found than a correlation. For 0–1 data for instance, simple measures can be derived from a 2 x 2 table of joint results (0, 0), (0, 1), (1, 0) and (1, 1) for the pairs of data. An appropriate test for independence would be Fisher's exact test.

This report is an official deliverable of the Welfare Quality® project.

Welfare Quality® is a European research project focusing on the integration of animal welfare in the food quality chain: from public concern to improved welfare and transparent quality. Welfare Quality® is co-financed by the European Commission, within the 6th Framework Programme, contract No. FOOD-CT-2004-506508.

Project Office Welfare Quality
Animal Sciences Group of Wageningen UR
Edelhertweg 15
P.O Box 65
NL - 8200 AB Lelystad
The Netherlands
Phone +31 320 293503
Fax +31 320 238050
e-mail: info@welfarequality.net
Internet: www.welfarequality.net

ISBN 1-902647-81-5
ISSN 1749-5164