Welfare® Quality

Science and society improving animal welfare

Assessment of Animal Welfare Measures for Layers and Broilers

edited by

B. Forkman and L. Keeling

SIXTH FRAMEWORK PROGRAMME

WELFARE QUALITY REPORTS NO. 9

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.

© Copyright is held by the authors of the individual contributions.

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

ISBN 1-902647-79-3
ISSN 1749-5164

CONTENTS

	Preface	*iii*
	List of Contributors	*vii*
1	Assessment of Water and Food Deprivation in Poultry	1
2	Standardisation of Clinical Scoring in Poultry	7
3	Standardisation of Measures of Broiler Lameness	31
4	Measures of Poultry Health Status	39
5	Assessment of 'Dead on Arrival' in Poultry	63
6	Assessment of Pre-stunning Shock and Stunning Effectiveness	67
7	Assessment of Human–Animal Relationship in Laying Hens	75
8	Assessment of General Fearfulness	91
9	Measuring Distress Vocalisation in Poultry	95
10	Positive Emotions in Poultry	103
11	Qualitative Behaviour Assessment	113
	Bibliography	121
	Appendix 1	137
	Appendix 2	153
	Appendix 3	167

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 poultry (layers and broilers). The measures included in this book have been assessed for adult layers (but not for layer chickens or roosters), and for broilers (but not broiler breeders).

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

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

protocols (e.g. for gait score in broilers), in these cases the project groups evaluated the alternatives and selected the most appropriate measure to be used.

This book is based on the work of different research groups, each chapter therefore deals with related aspects of animal welfare, e.g. 'clinical measures', related to various aspects of welfare, e.g. feather pecking, keel bone damage, etc. This means that although there are only 10 groups of measures/chapters, the number of measures evaluated is much higher since each group contains information on a number of different measurements. The measure 'Dead on arrival' (Chapter 10) is affected by a very large number of variables, and has been assigned to general fear for convenience.

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. These have however not been tested in the current project. Finally there is a third appendix with a brief note on statistics and sampling.

A consideration when working with poultry is the risk of causing panic among the animals. Any thorough inspection of individual animals means having to catch them, and any method for catching the birds has a high risk of causing panic and so called clumping, especially in layers. Many criteria therefore have to be assessed at slaughter rather than on farm.

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

1. Absence of prolonged hunger. Feed content of the muscle stomach at slaughter was investigated as a possible measure for feed restriction around this time (Chapter 1). On farm assessment of prolonged hunger may probably be done using resource based measurements based on access to feeders (resource measures appendix). For broilers access to weight data can be used (management measures appendix).
2. Absence of prolonged thirst. Hematocrit level at slaughter was investigated as a possible measure for thirst around (Chapter 1). For on farm measurement access to drinkers are a possibility (resource measures appendix).
3. Comfort around resting. Especially for broilers the litter quality is of high importance, this can be assessed as a resource based measure. Animal based measures which can be used and for which there are protocols are feet conditions (including hock burns), breast blisters and body cleanliness, and for layers keel bone deformations. For broilers can probably best be done at the slaughter plant (Chapter 2).
4. Thermal comfort. No animal based measures were investigated for this criteria. A possibility is to measure it using a resource based measure; the temperature at different sites in the house.
5. Ease of movement. This may be measured as the stocking density (Appendix 2).
6. Absence of injuries. This is most easily measured as carcass damage on plant at the inspection point immediately after de-feathering: skin damage; foot and toe damage;

and, limb fractures/dislocations (Chapter 2), but may for layers also be done on farm. For broilers an important measure is the gait score (Chapter 3).

7. Absence of diseases. The diseases can most easily be assessed at slaughter. The measurements include, but are not limited to, frequencies of enlarged crop, ascites, dermatitis/cellulitis, hepatitis, jaundice, pericarditis, septicae-mia, and enteritis (Chapter 4).The number of animals that are dead on arrival obviously depends on a large number of different variables, e.g. thermal comfort during transport and disease or injuries (Chapter 5).
8. Absence of pain induced by management procedures. The relevant management procedure is beak trimming (in layers), the beak trimming can be done at different ages and in different ways all of which affect the bird in known ways (Appendix 2). As for the other species the quality of stunning at slaughter is also relevant (Chapter 6).
9. Expression of social behaviour. The only relevant social behaviour recorded for domestic chickens are aggression, darting and sparring. The latter two only occur in very young birds, i.e. are only expected in broilers and not adult layers, and no protocol was developed to study these. Aggression is evaluated as its consequence, in this case wounds on comb and wattles (Chapter 2).
10. Expression of other behaviours. The expression of 'other behaviours' covers all behaviour that is not covered by other criteria. No stud-ies were done on this criterion.
11. Good human animal relations. The avoidance distance between human and animal is often used as a measure of human–animal relations. This measure was investigated here (chapter 7). This is also taken up in the management appendices.
12. Positive emotions. This criteria covers both negative and positive emotions. There are various fear tests for poultry such as the open field or tonic immobility test. For on farm assessment the novel object test was considered the most feasible (Chapter 8). Distress might also be assessed by the spontaneous behaviour of the animal, such as distress vocalisations (Chapter 9). The assessment of positive emotions is evaluated in Chapter 10. Finally, both negative as well as positive emotions might be assessed by qualitative behaviour assessment (Chapter 11).

B. Forkman
September 2009

LIST OF CONTRIBUTORS

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

C. Arnould
INRA, Nouzilly, France

A. Butterworth
University of Bristol, United Kingdom

B. Engels
Wageningen University and Research Centre, Netherlands

B. Forkman
University of Copenhagen, Denmark

E. Fröhlich
Zentrum für tiergerechte Haltung Geflügel und Kaninchen, Zollikofen, Switzerland

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

U. Knierim
University of Kassel, Germany

T. Lentfer
University of Kassel, Germany

K. Niebuhr
University of Veterinary Medicine Vienna, Austria

L. Nordensten
Swedish University of Agricultural Sciences, Uppsala, Sweden

M. Raj
University of Bristol, United Kingdom

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

V. Sandilands
Scottish Agricultural College, Edinburgh, United Kingdom

H. Schulze Westerath
University of Kassel, Germany

M. Staack
University of Kassel, Germany

A. Velarde
IRTA, Girona, Spain

S. Waiblinger
University of Veterinary Medicine Vienna, Austria

F. Wemelsfelder
Scottish Agricultural College, Edinburgh, United Kingdom

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

P. Zimmerman
Noldus Information Technology, Netherlands

1

ASSESSMENT OF WATER AND FOOD DEPRIVATION IN POULTRY

B. Algers and R. Westin

1.1 SUMMARY

Two field experiments were conducted to assess if hematocrit and feed content of the muscle stomach are reliable indicators of short-term water and food deprivation in slaughtered poultry. Sampling were conducted at 0, 4, 6, 8, 10, 12, 16, 20 and 24 hours after feed/water withdrawal. For measurement of hematocrit, individual blood samples were collected from 10 38-day old broilers at each sampling time point. Hematocrit was analysed using a micro hematocrit capillary tube reader. The feed content in the muscle stomach was scored for 5 birds at each sampling time point. Scoring was done by visual inspection of the level of feed content in the dissected muscle stomach using the scale 1 (low), 2 (medium) and 3 (high). Every muscle stomach was documented on photo. In addition to the field experiments, the feed content from 10 broilers from each of 8 different flocks slaughtered at an abattoir were scored and documented on photo. Feed withdrawal times extended from approximately 8.5–13.5 hours.

Statistical analysis of hematocrit levels showed that no rise in hematocrit was found during 24h of water withdrawal except in female birds at 20 hours of water deprivation. Hematocrit levels are hence not regarded as a reliable parameter for estimation of short-term water deprivation in poultry.

Scoring results from the experiment of food deprivation, suggest that a decrease in feed content is possible to measure after 10 hours of feed withdrawal but it does not seem to be linear over time up to 24 hours. The lowest median score were obtained in birds deprived of food for 16 hours. All sampled flocks at the visited abattoir had a mean score of 3, indicating that no decrease in feed content was seen in spite of up to 13.5 hours of food deprivation. In our opinion, this makes scoring of muscle stomach content an unreliable indicator of short term food deprivation in slaughtered poultry.

1.2 INTRODUCTION

Live animal transport from farm to processing plant is of great concern regarding to the welfare of transported animals. Catching and transportation of poultry induce stress (Nijdam et al., 2005) and it is therefore of high interest to keep the time period from loading at the farm to slaughter at the abattoir as short as possible to increase the animal welfare. A survey at four abattoirs in the UK has shown that the average total marketing time, from start of loading at a farm to the completion of unloading at the processing plant, for broiler chickens was 3,6 hours (Warris et al., 1990). However, for 6 % of all birds this time exceeded 7 hours and the maximum recorded total marketing time was 12,8 hours. Increasing transport and lairage time results in increasing numbers of birds dead on arrival (DOAs) (Nijdam et al., 2004). Today there are no animal based measures in use for estimation of the duration of time from loading to slaughter. This means that it is not possible to investigate if transport and lairage time exceeds its legal limits by checking the animals. Since most birds have no access to water during transport and lairage, hematocrit as an indicator of dehydration, is proposed as a feasible animal based measure. Hematocrit measures the proportion of red blood cells (erythrocytes) in the blood (Sturkie, 1986). When the plasma volume decreases due to low water intake, hematocrit rises. Knowles et al. (1985) found an 4.5 % increase of hematocrit in broiler chickens deprived of water for 24 hours. Dehydration could be an important factor prejudicing the welfare of transported poultry (Warris et al., 1993). A measure of food deprivation is also desirable for estimation of the time period from loading to slaughter. Content of feed in the muscle stomach is proposed since the crop content is highly variable and the content of the intestine does not very well reflect the feed consumption within the range of 24 hours.

The objective of this study was to assess whether hematocrit and feed content in the muscle stomach are reliable indicators of short-term water and food deprivation in slaughtered poultry. Sampling methods must be simple and possible to conduct on an abattoir.

1.3 METHODS

The study was carried out on a farm in Sweden. 38-day-old broilers of mixed sex were used. Before the study the birds were conventionally reared on sawdust litter in a building section containing a flock of 17 250 broilers. Commercially compounded feed mixed with whole wheat was available ad libitum from round feeding troughs and water from nipple drinkers with drip cups. At the start of the study, 90 birds where selected by convenience sampling and separated from the rest of the flock and put in a pen (325 x 390cm) without any water supply (trial 1). Another 45 birds where selected similarly on the same day and

put in a pen (120 x 390cm) without any feeding troughs (trial 2). Both pens where placed in the same building section as the original flock. At the day of the study the room temperature were 20ºC on average and during the 24h-period studied the birds spent 2 hours in darkness (00.00–02.00h). Trial 1 started at 08.30h and trial 2 at 13.00h. Blood samples from 10 water deprivated birds, were collected from the wing vein at 0, 4, 6, 8, 10, 12, 16, 20 and 24 hours after onset. New individuals were sampled at every time point. The birds were also individually weighed and sex was noted. Blood samples were immediately centrifuged in a micro-hematocrit centrifuge (StatSpin III) and hematocrit was analyzed (Micro hematocrit capillary tube reader, StatSpin® Technologies, Norwood, USA). One sample was lost due to technical failure.

Five food deprived birds were euthanized with the same time intervals as blood sampling in trial 1 (0,4,6…24h). Their muscle stomachs were dissected and then photographed with a digital camera. The photos were visually grouped and a scale for scoring of feed content level was produced were a score of 1) indicates a low level of feed content, 2) a medium level and 3) a high level. During a visit at an abattoir, muscle stomachs from 10 birds within each of 8 flocks were collected and scored with the produced scale. The shortest time interval from loading at the farm until start of slaughter was 2,5 hours and the longest 7,5 hours. To reduce carcass losses due to contamination of faeces, all flocks had been deprived of food for approximately 6 hours before loading and transport to the slaughterhouse began. This made the flocks represent a time interval of food deprivation from 8.5–13.5 hours.

Statistical analysis of hematocrit was carried out by multiple regression in the JMP Statistical Discovery software (version 6.0) and the GLM procedure of the SAS Statistical Software package (version 9.1). Hematocrit (log-transformed, to approximate Normality) was regressed on predictors representing number of hours of water deprivation, sex of the bird (1=male; 2=female) and body weight in kg (continuous). The model was developed in a manual step-wise procedure, including and excluding possible predictors to maximize the adjusted R-square. The least squares means for log hematocrit were used to compare non-deprived (0 hours) and deprived birds at different time points (4, 6, 8… 24 hours) by F tests, without adjustment for multiple comparisons. The models were validated by visual examination of residual plots for Normality, variation and outliers.

The photos of the food deprived birds from the field trial and the abattoir were studied and compared visually. No statistical analysis was carried out.

1.4 RESULTS

Table 1.1 shows the hematocrit in sampled birds of each gender at different times of water deprivation, as predicted by the statistical model. According to the model there was a weak association between hematocrit and water deprivation at 20 hours in female birds (deprived

birds had a 13% higher hematocrit than non-deprived; P=0.043) but not in males or at any other time point.

Scoring results of feed content at the field trial of food deprived birds are shown in Table 1.2. After 10 hours of feed withdrawal decrease in feed content was seen but decrease were not linear over time. The lowest median score were obtained in birds deprived of food for 16 hours.

The scoring results of feed content from birds slaughtered at the visited abattoir is shown in table 3.

No obvious difference in food content level was found between the flocks. All flocks had a median score of 3.

1.5 DISCUSSION

The results show that there is almost no change in hematocrit during 24 hours of water deprivation in poultry during farm conditions. The only statistical significant difference found was in females deprived of water for 20 hours. 24 hours after onset of the trial there was no difference. This is in contrast to results by Knowles et al. (1995) who found an 4,5 % increase of hematocrit in broiler chickens deprived of water for 24 hours. In their study, hematocrit was only measured at one time point (24-h) so it is not possible to tell if the rise was only temporary or not. The authors however regarded this as relatively small change, indicating that birds can maintain their blood volume very effectively during 24 hours of dehydration. This is also confirmed by Koike et al. (1983) who measured plasma volume and hematocrit among other things, in White Leghorn pullets dehydrated from 24 to 96 hours. Hematocrit levels were significantly elevated when dehydrated for 48 and 72 hours but not at 24 hours. A short-term study of dehydration (≤12h) also concludes that hematocrit levels are unchanged during the early stages of water deprivation in broiler chickens (Chamblee et al., 1989). To be able to estimate the time period from loading to slaughter by a blood sample we need a parameter that change, preferable linear over time, during early stages of water deprivation. This experiment clearly shows that this is not the case for hematocrit during 24 hours of dehydration. Our conclusion is therefore that hematocrit is not feasible to use in this purpose. Since we found that this parameter showed poor validity, no measurements of hematocrit on regular slaughtered birds were conducted as was proposed in the original plan.

The result from our field experiment of food deprivation shows that the feed content of the muscle stomach is not linear over time even if a decrease in food content was possible to detect after 10 hours of feed withdrawal. All sampled flocks at the visited abattoir had a mean score of 3, indicating that no change was seen in spite of up to 13.5 hours of food

TABLE 1.1 Hematocrit (%) in 89 broiler chickens deprived of water for different time periods (back-transformed least squares means).

Hours of water deprivation	Males	Females
0h	31.0 (n=5)	32.6 (n=5)
4h	29.8 (n=5)	32.5 (n=5)
6h	33.2 (n=5)	28.9 (n=5)
8h	32.0 (n=5)	31.0 (n=5)
10h	31.7 (n=4)	31.2 (n=5)
12h	31.5 (n=5)	31.6 (n=5)
16h	32.5 (n=5)	34.4 (n=5)
20h	31.2 (n=4)	37.1[a] (n=6)
24h	31.3 (n=5)	32.9 (n=5)

Note: [a] Mean differs significantly from mean at 0h within the same column (P≤0.05).

TABLE 1.2 Feed content in muscle stomach of 50 broiler chickens deprived of food for different time periods.

Hours of food deprivation	Number of birds within each score category			Median score
	Score 1 (low)	Score2 (medium)	Score 3 (high)	
0			5	3
4			5	3
6			5	3
8			5	3
10		3	2	2
12		4	1	2
16	3	1	1	1
20	1	3	1	2
24	1	2	2	2

TABLE 1.3 Feed content in muscle stomach of 80 slaughtered broiler chickens deprived of food for different time periods.

Flock	Hours of food deprivation	Number of birds within each score category			Median score
		Score 1 (low)	Score2 (medium)	Score 3 (high)	
1	13.5			10	3
2	13		1	9	3
3	13			10	3
4	11.5		2	8	3
5	11		2	8	3
6	10.5		4	6	3
7	10		1	9	3
8	8.5		1	9	3

deprivation. In our opinion, this makes scoring of muscle stomach content in slaughtered poultry an unreliable indicator of short term food deprivation. Slowdown of feed passage time occurs through limitation in movement of digesta from the crop (May et al., 1990). It is proved that confinement and darkness slowdown the passage of digesta from the crop (May et al., 1990). Transportation and lairage involve both factors. A common opinion of

farmers and people working at poultry slaughter plants is that birds do not digest during transport (Lindblad, personal message, 2006). All slaughtered birds in our study had a feed withdrawal time of approximately 6 hours before loading and 86% of all individuals had a high level of feed content in the muscle stomach. All birds deprived of food for 6 hours in the field trial had a muscle stomach filled with digesta. According to these findings, the feed content in the muscle stomach of slaughtered birds seems to reflect the feed withdrawal time at the farm rather than the total time of food deprivation including transport and lairage times.

According to the original plan, 10 (trial 1) respectively 5 (trial 2) birds were proposed to be sampled every hour from 4 until 12 hours of food/water deprivation. Since the time span from loading at the farm until slaughter can exceed 12 hours (Warris et al., 1990) it was decided to extend withdrawal times up to 24 hours.

1.6 CONCLUSIONS

Hematocrit and scoring of feed content in the muscle stomach are not valid measures of short term food and water withdrawal times in slaughtered poultry.

2

STANDARDISATION OF CLINICAL SCORING IN POULTRY

C. Arnould, A. Butterworth and U. Knierim

2.1 SUMMARY

On the basis of a literature review, our own practical and scientific experience and discussion between the collaborators, an on-farm assessment scheme for alterations of the integument and bones of broilers and laying hens is proposed that can be used in the framework of a broader welfare assessment tool. The scheme discussed considers six different measures taken on-farm or at slaughterhouse. Measures of body cleanliness, breast blisters and feet condition (pododermatitis, hock burns) are proposed for broilers. Measures of plumage condition, wounds, feet condition and bone condition (keel bone damage) are proposed for laying hens. In the published studies a wide variety of different methods regarding detail as well as regarding the type of the qualitative and quantitative descriptions of the alterations were used. Accordingly, either a severity scale or a determination of the presence or absence of the alteration are proposed for on-farm assessment. In broilers, all the measures proposed are performed at slaughterhouse except for body cleanliness, which may also be measured on farm. Three or four-point scale scoring systems are proposed, except in the case of breast blisters where presence – absence at assessment is proposed. This latter type of assessment is also proposed as an alternative for hock burns. In laying hens all measures proposed are performed on farm, even if in some cases measures at slaughterhouse can be carried out. For keel bone damage and pecks at comb, a presence–absence recording at assessment is proposed. A three-point scale scoring system is proposed for all other measures.

2.2 INTRODUCTION

The objective of this project was the standardisation of clinical scoring measures performed on farm and at slaughterhouse to evaluate broiler and layer welfare. Specifically,

it was the task to propose standardised methods of measurement that fulfil the criteria of validity, reliability and feasibility. Clinical scoring measurements concern:

- wounds;
- plumage condition;
- body cleanliness;
- breast blisters;
- feet condition including hock burns;
- broken bones including keel bones damages.

According to the measures involved, assessments are performed on the farm or at the slaughterhouse. Some of the measures concern only one type of bird: broilers or laying hens.

This work is based on a literature review on clinical scoring in broilers and laying hens. Proposals made in this report have been discussed by the collaborators of this project. With respect to clinical scoring in laying hens, scoring systems proposed in the Laywel project have been taken into account and the scoring system proposed in this report has been simplified to produce the scheme described here.

The condition of the plumage and skin, including the skin of the feet, reflect the ability of the birds to maintain their bodily integrity and to avoid harmful, and potentially painful, interaction with its environment. Moreover, plumage and skin condition may be an indicator of the quality of management as regards floor, litter or perch dryness and cleanliness, air quality and risk of injury from housing equipment. The condition of feathers and skin is, therefore, an important direct and indirect measure of animal welfare which has been used in a considerable number of studies. Plumage condition may be assessed with regard to cleanliness and completeness, the skin regarding wounds, breast blisters, but also at the feet regarding alterations such as discolouration, excessive growth or ulceration. To the occurrence of bone fractures apply similar considerations. It was the objective of this part of the task to review wounds, plumage condition, body cleanliness, breast blisters, feet condition including hock burns and bone condition including keel bones.

Wounds: Wounds are mainly a problem in laying hens, where cannibalism is sometimes observed. As far as we know this is not the case in broilers. Even if scratches are probably of importance in broilers, we are not able at the moment to make a proposal for a valid and feasible scoring system. However, it might be of interest to add a measure of scratches in the on-farm assessment and then some suggestions are given.

Plumage condition: At the moment, this parameter is more relevant for laying hens than for broilers and is then discussed with reference to laying hens.

Body cleanliness: This parameter concern both broilers and laying hens, but is most relevant for broilers at the moment. Then, no proposal is given for laying hens.

Breast blisters: At the moment, this parameter is more relevant for broilers than for laying hens and is then discussed with reference to broilers.

Feet condition including hock burns: These measures are mainly of concern in broilers, but feet condition also for laying hens. Pododermatitis and hock burns are most often seen in broilers, whereas hyperkeratosis and bumble foot is seen in laying hens. Furthermore, broilers often have dirty feet but laying hens do not and then their feet usually can be well assessed on farm. The assessment therefore differs between broilers and laying hens in the scoring system used and in the location were measure should be done (at slaughterhouse versus on farm).

Broken bones including keel bones damages: These measures concern mainly laying hens. No proposal is given for broilers.

2.2.1 Note

The scoring systems proposed in broilers and laying hens are based on a 'low score' direction – in which a low score means low damage and a high score means high damage. This is in contrast (reverse direction) to what is proposed by Tauson in the EU-Laywel project (i.e. a low score means high damage and a high score means low damage) but is consistent with the overall Welfare Quality® approach for measures, where a low score indicates a 'better welfare outcome'.

2.3 BROILERS

2.3.1 Introduction

This part reviews wounds, plumage condition, body cleanliness, breast blisters, feet condition (pododermatitis and hock burns) and broken bones.

In most of the studies performed on broilers, clinical scoring has been performed at slaughterhouse. However, for some of the measure an assessment on farm is possible.

2.3.2 Wounds

Validity

In broilers very few authors have looked at wounds on farm or at slaughterhouse. It is probably due to the fact that wounds are rare in broilers, except in the case of scratches when observations are performed at slaughter house after transportation. In the past it seems that scratches were rather rare on farm (Bruder, 1988) but it needs to be confirmed. Measurement of scratches on slaughterhouse line has not been validated. According to the study of Colas (2005) a lot of birds have scratches when all types of scratches (slight and severe) are observed at slaughter line and the assessment of scratches is obviously of importance in a welfare point of view. The method of assessment used in this study does not seem to be highly discriminating, but an evaluation of the birds with severe old scratches or a high number of superficial old scratches on the back as proposed by Elfadil et al. (1996) could be a valid measure of the incidence of scratches on farm. According to Frankenhuis et al. (1991) a decrease in density and an increase in feeding place result in a reduction of scabby hips due to scratches at slaughterhouse. Bruises are observed on slaughter line but it seems quite difficult to differentiate easily between those appearing before and after the electrical stunning. It is impossible to assess the cause of bruise damage and visual examination can not be used for bruises of less than 12 hours of age (for a review, see May and Hamdy, 1966).

Proposal

Due to the considerations exposed above, we therefore suggest that the presence/absence of wounds is not measured for broilers. However, wounds due to feather pecking and cannibalism (see laying hens) could concern low growth rate meat chickens in some cases.

2.3.3 Plumage Condition

Validity

At the moment, the assessment of plumage condition is relevant for laying hens only. However, it could concern low growth rate meat chickens in some cases. A scoring system derived from Tauson et al. (1984) and used on laying hens (Guesdon, 2004) has already been used on low growth rate broilers (Chapuis et al., 2003).

2.3.4 Body Cleanliness

Validity

Maintenance of the plumage in terms of cleanliness is important for thermoregulation. It could, moreover, be an indicator of the quality of the management. Body cleanliness has been evaluated on farm in broilers and laying hens but very few studies have taken this measure into account. Very few authors have looked at leg or feet cleanliness but as plumage cleanliness it could be an indicator of the quality of the management.

Measure should be done on farm as effect of transport on cleanliness could be of importance. Body could become dirty during transport as birds are put in crates in high density for several hours but in this case body cleanliness does not reveal welfare on farm.

The validity of body cleanliness to assess welfare in broilers has not been adequately evaluated. However, we can assume that it may be a good indicator of the quality of management, pre-slaughter starvation time for birds assessed at slaughter (Wilkins et al., 2003) and litter quality (humidity).

Assessment schemes used in different studies

Body cleanliness is often assessed using scoring systems (Akpobome and Fanguy, 1992; Weeks et al., 1994; Elwinger, 1995) but details on the method used are often missing. Elwinger (1995) assess cleanliness of the plumage (score 1 to 5) and of the feet as to the occurrence of dirty lumps attached to the claws (score 1 to 3) in broilers. Overall dirtiness including that of the legs has been assessed in broilers by Weeks et al. (1994), their scoring system (score 0 to 2) taking into account surface of soiling and the presence of caked dirt and dirt adhering the feathers. Dawkins et al. (2004) have registered dirty pads in broilers but the method is not described and it is not clear if the measure was performed on farm or at slaughterhouse.

Elwinger (1995) observed in broilers that the percentage of dirty feathers varied between seasons and farms, but was not influenced by stocking density. Feet cleanliness was affected by these three factors. Weeks et al. (1994) observed that dirtiness is negatively affected by the density and that birds with an access to outdoors were cleaner than broilers reared indoors.

Wilkins et al. (2003) describes an 8-point photographic scale for broiler bird cleanliness used to assess the hygiene of the plumage in birds arriving at slaughter (Figure 2.1). This study links bird cleanliness with feed withdrawal times and litter and housing hygiene conditions. Dirty birds may be created by poor transportation conditions and this may be difficult to separate from housing conditions

FIGURE 2.1 Broiler body cleanliness scoring scheme.
Source: Wilkins et al., 2003; reproduced with permission of the author.

Feasibility

The measure of body cleanliness is easy to perform but needs handling of the birds on farm. It is readily possible pre slaughter using the method of Wilkins et al. (2003).

Proposal

For measure on farm, we propose a three-point scale scoring system derived (simplified) from Weeks et al. (1994):

Score 0: Feathers white, no caked dirt on legs and feet
Score 1: Either moderate soiling all over body or variable soiling with no more than half the body or legs having caked dirt and most feathers free
Score 2: Most of body and feet caked with dirt adhering the feathers to each other.

For measure at slaughterhouse, we propose a four-point scale scoring system derived (simplified) from 8-point scale from Wilkins et al. (2003):

Score 1 (very clean) regrouping Wilkins' scores 1 and 2
Score 2 regrouping Wilkins' scores 3 and 4
Score 3 regrouping Wilkins' scores 5 and 6
Score 4 (very dirt) regrouping Wilkins' scores 7 and 8.

Good light is required to score the birds and this may run counter to the general trend to have subdued lighting in lairage and hang on areas in poultry slaughter plants. Furthermore, birds may become dirty in the crates after their removal from the farm.

2.3.5 Breast Blisters

Validity

Breast blisters are clearly a valid parameter in term of welfare. They represent discomfort and the potential for damage to the skin resulting in chronic infection. The relationship with feet condition is not clear (Colas, 2005; Mirabito et al., 2006) and this is why breast blisters needs to be included in the welfare assessment scheme. From a preliminary experiment it seems that genotype could influence the prevalence of this lesion (Colas, 2005).

Feasibility

Measure of breast blisters is easy to perform on the slaughter line but needs handling of the birds on farm.

Assessment Schemes Used in Different Studies

We found only one author measuring breast blisters on farm (Elwinger, 1995). Breast blisters were scored from 1 (the best) to 5 (the worst), but no definition of these scores is given. Measures of breast blisters are usually performed at the slaughterhouse because it is too difficult to perform on farm due to dirtiness of breast region. Few references exist on breast blister measurement at slaughterhouse in broilers. Usually, methods used to assess breast blisters are not fully described (see for example Martrenchar et al., 1997; Hall, 2001). However, in some cases scoring systems using a three- to four- point scale have been used, but there is no clear definition of the scores (Stephenson et al., 1960; Pettite et al., 1983 – broiler breeder hens). Colas (2005) has used a determination of the presence or absence of the skin alteration corresponding to a breast blister of more than or equal to 0.5 cm^2. Examples of moderate and severe breast blisters that could be scored as 'present' are given in Figure 2.2.

Proposal

Colas (2005) presence–absence evaluation could be used at the slaughterhouse (see above). An alternative could be to create a four-point scale (no lesion, mild, moderate, severe) but the reference (specific description of these score) would need to be created.

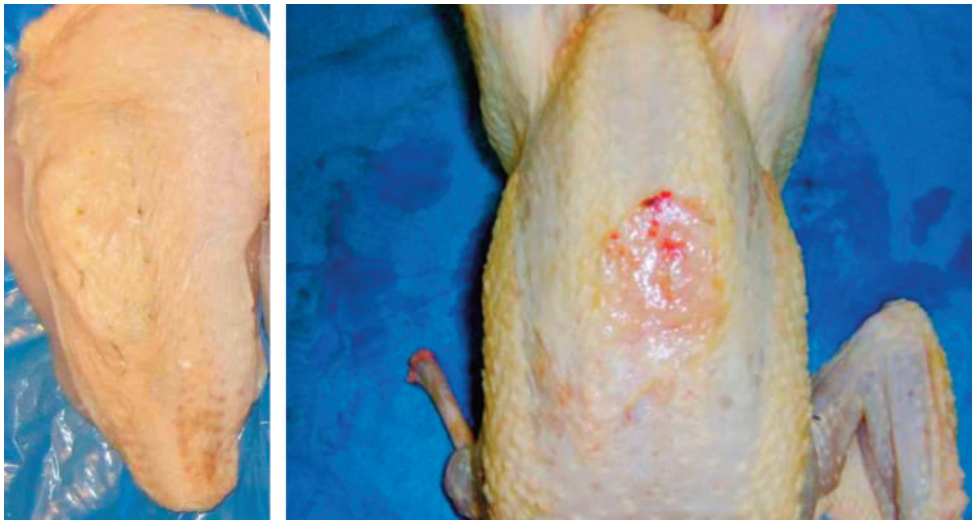


FIGURE 2.2 Example of mild breast blister without ulceration and severe breast blister with ulceration.
Source: A. Butterworth, University of Bristol (photo).

2.3.6 FEET CONDITION (PODODERMATITIS AND HOCK BURNS)

Validity of Pododermatitis Measures

This parameter is clearly valid and should be included in the welfare assessment scheme. It will represent discomfort and the potential for damage to the skin resulting in chronic infection.

Pododermatitis is a direct cause of lameness (Kestin et al., 1999; Martrenchar et al., 2002) and can probably induce suffering in the case of severe lesions (Ekstrand et al., 1997). Pododermatitis is linked with litter quality, litter humidity, genotype and by factors affecting litter quality such as diet, drinker design, density, season, climate control (for review Mayne, 2005). A measure of pododermatitis is used in Sweden to control welfare of broilers and this measure was included in the project of Directive for the protection of chickens kept for meat production (2005).[1]

Assessment Schemes Used in Different Studies for Pododermatitis[2]

Most of the measures related to feet condition (pododermatitis and hock burns) in broilers are performed at the slaughterhouse as removal of crusted faeces and handling of the birds

[1] The Council Directive laying down minimum rules for the protection of chickens kept for meat production (Council Directive 2007/43/CE) is now adopted (since the 28 June 2007). In Annex III it is writen: 'In the context of the controls performed under the Regulation (EC) No 854/2004, the official veterinarian shall evaluate the results of the post-mortem inspection to identify other possible indications of poor welfare conditions such as abnormal levels of contact dermatitis…'.
[2] Several papers other than those mentioned in this paragraph have been published recently. They have not been taken into account in this report, which was delivered in 2006.

is necessary to assess this on farm. Pododermatitis has been assessed on farm in live birds in few studies and scoring systems used were different. Furthermore, score definitions are not clear. Mc Lean et al. (2002) has measured contact dermatitis of the feet, taking into account 'burning' and scab (score from 0 to 4). Elwinger (1995) scored feet pads according to the occurrence of injuries (scores from 1 to 2 with 1 is best). Dawkins et al. (2004) scored pad dermatitis (score from 0 to 2) taking the surface area of the lesion into account but they did not specify if the measure was performed on farm or at the slaughterhouse.

In most of the studies, pododermatitis is measured at slaughterhouse. Many different methods have been used to assess pododermatitis. Scoring systems varied in the number of scores used: from a three-point scale (Harms et al., 1977) to an 11 point scale (Wang et al., 1998). Scoring systems used are based on the severity (Ekstrand et al., 1998) or on the absolute (Dawkins et al., 2004) or relative (Martrenchar et al., 2002) surface area of the lesion. Sometimes severity and surface are combined (Clark et al., 2002; Colas, 2005; Mirabito et al., 2006). Colas (2005) proposed a simplified three-point scale scoring system that would be more efficient if a lot of birds have to be scored at slaughter line. This scale needs to be further validated. Pagazaurtundua and Warriss (2006) have recently used a scoring system based on the severity of the lesions of footpad dermatitis and observed a linear relationship between the square root of the relative area of the lesions and their scores. Their scoring system contained a four-point scale and they observed a very high correlation between two assessors with this scale (r=0.998). Figures 2.3 and 2.4 present the scoring systems used by government in the United Kingdom and Sweden: foot burn scale (FBS) and the Swedish foot pad lesion classification (SFPLC).

Feasibility for Pododermatitis

Measure of pododermatitis is easy to perform at slaughterhouse.

Proposal for Pododermatitis

Taking the scoring system presented above (FBS, SFPLC; Figures 2.3 and 2.4) and the results of Colas (2005)1 and Mirabito et al. (2006) into account, we propose the scoring system below.

Score 1 corresponding to score 1 and 2 of FBS with lesion < 25% of the foot pad.
Score 2 corresponding to score 3 of FBS
Score 3 corresponding to score 4 and 5 of FBS

This scoring system will be in accordance with the use of three-point scale scoring system as it is proposed in the project of Directive (2005)2 and probably will increase the inter-assessor repeatability as the number of scores is lower and the definitions given for the scores are more precise.

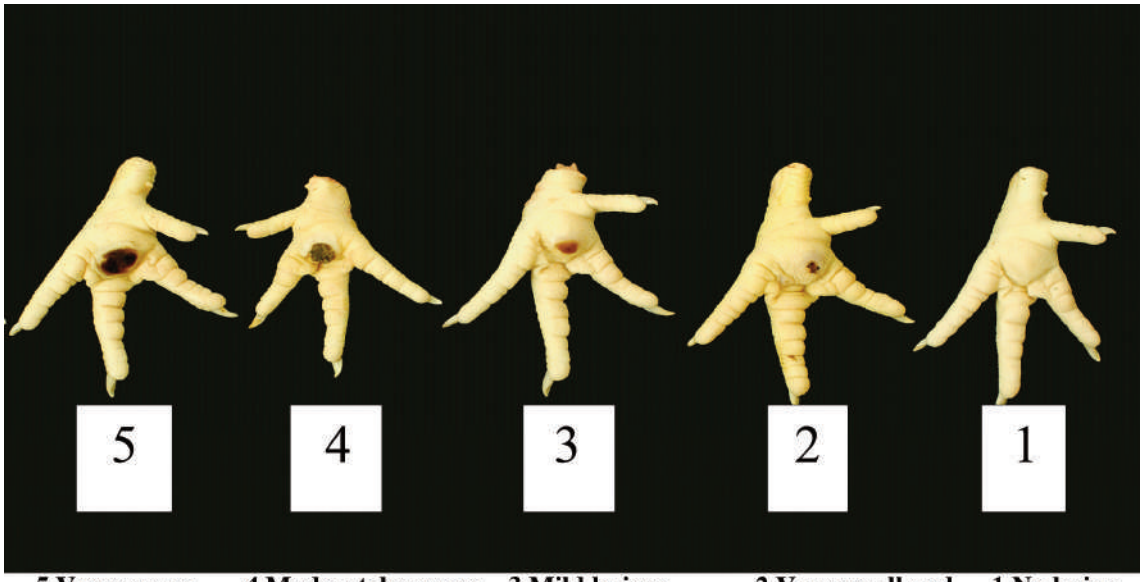


Figure 2.3 Scoring systems used in the United-Kingdom (foot burn scale, FBS).
Source: Department for Environmental Food and Rural Affair, UK.

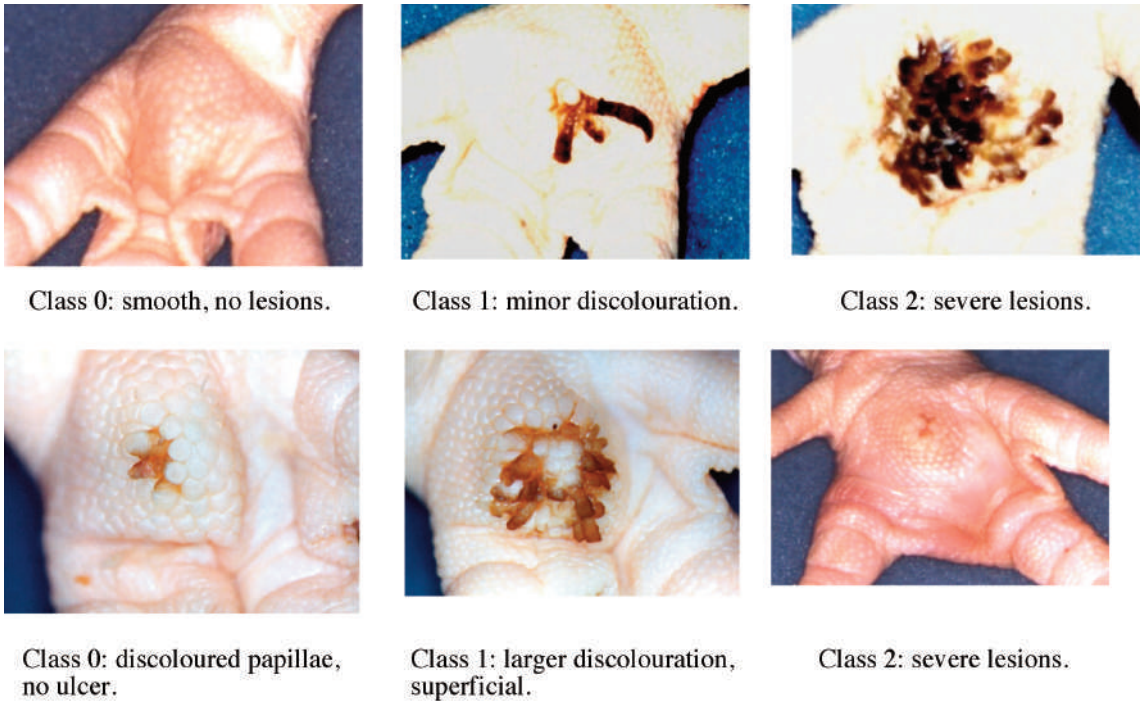


FIGURE 2.4 Scoring system used in Sweden (Swedish foot pad lesion classification, SFPLC).
Source: Department for Environmental Food and Rural Affair, UK.

However, recently Pagazaurtundua and Warriss (2006) used a four-point scale with the first score corresponding to 'no lesion'. Then, the same principle can be proposed, i.e.:

Score 1 corresponding to score 1 of FBS, i.e. no lesion;
Score 2 corresponding to score 2 of FBS with lesion < 25% of the foot pad;
Score 3 corresponding to score 3 of FBS;
Score 4 corresponding to score 4 and 5 of FBS.

Validity of Hock Burn Measures

Hock burn is clearly a valid parameter in terms of welfare. Hock burns induce discomfort and the potential for damage to the skin resulting in chronic infection. Some authors have found a relationship between the occurrence of hock burns and foot pad dermatitis, hock burns appearing later than pododermatitis (Greene et al., 1985; Colas, 2005)1. The welfare of a bird affected by pododermatitis and hock burn is probably worse than that of a bird affected only by pododermatitis. In the UK 2006 survey carried out for DEFRA (Anon, 2006), each increment in hock burn score was associated with an increase in gait score of 0.27. Hock burn may be both causal and associative – it may contribute to lameness but could equally, wholly or in part, be caused by more time spent sitting on the litter. Risk factors are the same than for pododermatitis. We believe that the measure of hock burns needs to be included in the on-farm welfare assessment scheme, however the reliability of the scoring system should be evaluated.

Assessment Schemes Used in Different Studies for Hock Burn

We can find only one author reporting the assessment of hock burns on farm. Contact dermatitis of hocks is usually measured at the slaughterhouse. Scoring systems used at slaughter vary between authors. Mc Lean et al. (2002), took into account 'burning' and scab (on a five-point scale, as shown in the Gleadthorpe scale; Figure 2.5). Limits are the same as for pododermatitis. Sørensen et al. (2000) and Kristensen et al. (2006) used a four-point scale scoring system but gave no clear definition of the scoring systems. Colas (2005)1 used a 7-point scale system, taking severity and surface of the damage into account. This preliminary study suggests the importance of the severity of the damage rather than the surface. It is also suggested that a presence – absence classification could be used with presence corresponding to visible discoloration on slaughter line, even if small or superficial. Dawkins et al. (2004) assessed hock burns (score from 0 to 2) taking surface of the lesion and discoloration into account, but as for pododermatitis it is not specified if the measure was performed on farm or at slaughter.

Feasibility for Hock Burn

Scoring of hock burns is easy to perform on slaughter line.

Proposal for Hock Burn

The Gleadthorpe scale (GHBS) presented above could be used for welfare assessment. However, we think that this scoring probably should be simplified if we want to have a good correlation between assessors. We propose a simplified scoring system based on the Gleadthorpe scale with only three-point scale:

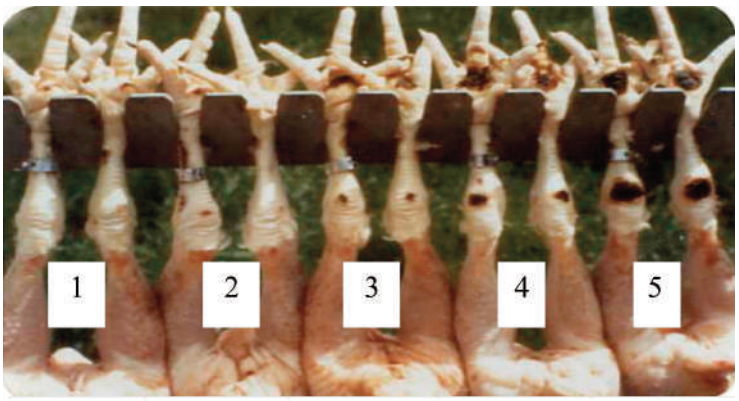
Score 1 corresponding to score 1 of GHBS;
Score 2 corresponding to score 2 and 3 of GHBS;
Score 3 corresponding to score 4 and 5 of GHBS.

This scale probably needs an evaluation for inter-assessors repeatability. An assessment using a presence – absence system could also be used as suggested by Colas (20051, see above).

In some abattoirs this measure would probably be difficult to carry out as legs are cut at the hock joint very early in the process.

2.3.7 Broken Bones

In broilers, broken bones (but above all dislocation of the hips) are mainly due to catching and hanging on slaughter line (Gregory and Austin, 1992; Sparrey and Kettlewell, 1994). Broken bones are clearly of importance from a welfare point of view. However, measure of broken bones on farm is not pertinent as no broken bone is observed in broilers and it



1 No lesion.
2 Very small and superficial lesions: slight discolouration on a limited area, mild hyperkeratosis.
3 Mild lesion: discolouration of the foot pad, superficial lesion, superficial dermatitis.
4 Moderately severe lesion: ulcers or scabs.
5 Very severe lesion: ulcers or scabs, signs of haemorrhages or deep dermatitis.

Figure 2.5 The Gleadthorpe hock burn scale (GHBS).
Source: Department for Environmental Food and Rural Affair, UK.

is not possible to measure at the slaughter line because it needs dissection or palpation. This is why we suggest excluding broken (leg) bones measure from the welfare assessment scheme. Keel bone damages are not a problem in broilers.

2.4 LAYING HENS

2.4.1 Introduction

This part review plumage condition (completeness), body cleanliness, wounds, feet condition and keel bone condition. The presence of overgrown claws and the condition of the beak (degree of trimming) are not considered here. While overgrown claws in cages present a welfare risk, the decision was taken that the protocol must be limited and that only the most welfare relevant measures will be included. For the degree of beak trimming we expect that this is taken into account within the management based measures.

In most of the studies performed on laying hens, clinical scoring has been performed on farm. However, for some of the measures an assessment at slaughterhouse, as it is done for broilers, is also possible.

2.4.2 Plumage Condition

Validity

Any plumage damage results in a welfare impairment in terms of the bodily state of the animal. Additionally, damage due to feather pecking indicates impaired welfare on several further lines: Feather removal has been associated with pain (Gentle and Hunter, 1990) and skin without feathers has a greater risk of injury. Moreover, the pulling out of flight feathers at the tail and wings by conspecifics may cause skin damage and lead to cannibalism, resulting in wounds at these locations. Birds with partial plumage loss face increasing demands on their metabolic rate in moderately temperated or cold environments (Deschutter and Leeson, 1986; Tauson and Svensson, 1980). Plumage abrasion indicates repeated collisions with housing equipment (Simonsen et al., 1980; Freire et al., 1999) or other birds, which might be perceived as aversive, although no experimental evidence is available on this. Feather damage may also be due to nutritional imbalance (e.g. lack of methionin or zinc) or intake of e.g. mycotoxins (Deschutter and Leeson, 1986) which again may negatively affect the birds physically and emotionally. Therefore, plumage condition is linked to feather pecking as well as housing conditions and feeding. It appears

to be the most feasible, reliable and valid measure out of these. However, in beak-trimmed birds the plumage condition might not truly reflect the actual degree of feather pecking behaviour.

When assessing feather damage it is to some extent possible to differentiate causes such as feather pecking, aggressive pecking or abrasion by the location of the damage. Poor feather cover on the head probably reflects that a bird has received aggressive pecks whereas feather damage on the belly, back, legs and neck probably reflects feather pecking. However, damage at the throat may well result from abrasion at the feed trough. Additionally, feather loss due to feather pecking at the breast is often difficult to differentiate from natural feather loss at the brood patch (Figure 2.6).

Assessment Schemes Used in Different Studies

There are numerous published scoring schemes that differ in many aspects. The plumage is often evaluated using numerical scores on a scale ranging form three-points (e.g. Gunnarsson et al., 2000) to 9-points (e.g. Bestman and Wagenaar, 2003). Most systems, however, use a four- or five-point scale. While in most schemes feather condition is the worse the higher score, in the schemes of Tauson et al. (1984, 2005) the higher the score the better is the plumage.

For the quantification of feather damage some studies used standard circular areas, for example they distinguished between naked areas 'less than 5 cm diameter' and 'greater than 5 cm diameter' (Bilcik and Keeling, 1999; Gunnarsson et al., 1999; Häne et al., 2000). In other studies, proportional measures have been used. For example, Bilcik and Keeling (1999) used alternatively to the standard area 'less than 50% of evaluated body area' or 'greater than 50% of body area evaluated', whereas Herremans and Decuypere (1987) used 'under one quarter of feathers worn/abraded' to 'over three quarters of feathers worn/abraded'. More general characterisations have also been used, such as 'almost completely covered with feathers', 'larger naked areas', 'no or very small areas covered

FIGURE 2.6 Brood patch.
Source: M. Staack, University of Kassel (photo).

with feathers' (Tauson et al., 1984). Adams et al. (1978) and Tauson et al. (2005) do not provide any definition, but only photographic illustrations of the scores.

Wide variation also exists with regard to the localisations scored. Sometimes the body is scored as a whole (Hughes and Duncan, 1972; Adams et al., 1978), divided into 11 areas (Bilcik and Keeling, 1999), 10 areas (Gunnarsson et al., 2000), 8 areas (Vestergaard et al., 1993), 6 areas (Herremans and Decuypere, 1987, Tauson et al., 2005), 5 areas (Tauson et al., 1984; Häne et al., 2000, Keppler et al., 2001, Bright et al., 2006) or 4 areas (Allen and Perry, 1975). For a study on feather pecking, Gunnarsson et al. (1999) took only denuded areas at the back of hens into consideration, because damage at this site is almost certainly caused by feather pecking.

Further differences relate to the final measures used. While in most studies a total body score is calculated by summing up all region scores, in some the percentage of birds having various scores is also used (e.g. Kjaer and Sørensen, 2002) or the proportion of birds with specified, localised damage (Gunnarsson et al., 1999; Huber-Eicher and Sebö, 2001).

Reliability

From their evaluation of their scoring scheme Tauson et al. (1984) concluded that only data obtained by the same assessor could be used for comparisons between birds/flocks. Although they found high correlations between the total scores of two assessors (r=0.94), the scores allocated for wings and tails differed significantly between the assessors. Herremans and Decuypere (1987) compared three schemes (Adams et al., 1978; Tauson et al., 1984; Herremans and Decuypere, 1987) and concluded that different scoring schemes vary with regard to repeatability, although the total body scores in general showed high correlations between two assessors (r=0.92, 0.93, 0.98). Differences with regard to reliability were confirmed for four other scoring schemes (Bilcik and Keeling, 1999; Gunnarsson et al., 2000; Keppler et al, 2001; Damme, unpublished) by Keppler et al. (2005). Here the method of Gunnarsson et al. (2000) showed the best agreements.

The statistical methods used for such comparisons are not universally agreed on. For example, correlation analysis has been used (Adams et al., 1978; Tauson et al., 1984; Herremans and Decuypere, 1987) as well as analysis of variance (Tauson et al., 1984) and determination of the relative deviation between scorers (Herremans and Decuypere, 1987). Gunnarsson et al. (2000) have stated that these methods may not be appropriate, and used Cohen's kappa test and the prevalence-adjusted bias-adjusted kappa test (PABAK). They reached good agreements between two pairs of assessors (PABAK>0.6) and often excellent agreements (PABAK>0.8) for all body regions, except in one assessor pair for the rump part of the back and wing pinions (PABAK=0.52, moderate agreement) and the breast (PABAK=0.13, poor agreement).

Although photographic illustrations of the scores can be provided to improve reliability (see e.g. Tauson et al., 1984; Gunnarsson et al., 2000), the types of feather damage may differ profoundly between hens from cages and non-cage systems. It may, therefore, be

difficult to generalise from the hens from those pictures to ones originating from another housing system.

Another approach to improve reliability is to agree on assessments between two or more assessors (e.g. Hughes and Duncan, 1972; Bilcik and Keeling, 1999) or to use the mean score of several assessors (e.g. Adams et al., 1978 from three assessors).

Feasibility

For most scoring procedures described hens have to be captured and handled individually. Therefore, they are relatively time-consuming, and the assessment of a statistically significant number of birds and ensuring a non-biased sample may be problematic. Moreover, lifting the birds up, holding and turning them in order to allow all parts of the body to be visually inspected is disruptive, even though it is a non-invasive method. Disturbance can also be caused by catching birds from different places in the poultry house in order to ensure a random sample. Apart from possible welfare concerns, farmers may fear production losses. In the published studies the number of birds assessed was often relatively low. Nicol et al. (2003) scored 15 birds out of flocks of 5000. Tauson et al. (1984) scored birds from 3 cages (4 or 5 birds/cage) out of 12-14 cages, Weber et al. (2003) scored 25 hens from flocks of 1541 to 2110 hens every two months and Gunnarsson et al. (1999) scored 100 hens per flock with sizes of 225 to 9954 hens. According to Bright et al. (2006) industry standards (e.g. ADAS Gleadthorpe, UK) require ≥ 100 birds to be scored. However, from a statistical point of view, in order to be able to estimate the true prevalence in a flock with a precision of for example ± 5 %, sample sizes of up to 385 birds may be necessary.

A scanning of whole flocks with estimation of the percentage of birds with naked areas or damaged feathers was 20 to 30 times faster than the scoring of captured individuals, but showed only sufficient reliability between two assessors for birds that had naked areas (r=0.86) (Keppler et al., 2005). This study was performed in small flocks. Assessment of birds in large flocks will most likely be more difficult.

Huber-Eicher and Sebö (2001), Bestman and Wagenaar (2003) and Bright et al. (2006) inspected individual birds from a distance. Huber-Eicher and Sebö only scored damaged tail feathers and naked areas equal to 2 cm in diameter or larger as absent or present in 10 % of the birds of a flock. Bestman and Wagenaar (2003) and Bright et al. (2006) applied different scoring schemes at different numbers of body regions. Bright et al. (2006) found high correlations for total scores between two assessors (r=0.91) and the scoring from a distance and in picked up birds (r>0.92), although slightly more damage was noticed in the picked up birds.

Proposal

Within the framework of the EU-Laywel project a proposal was made for a scoring system including 6 body regions and a four-point scale (Tauson et al., 2005). We agree that for experimental purposes this system may be appropriate, but for welfare assessment we propose to use the simplest combination of the systems reviewed. This means that the assessment should be carried out using a three-point scale and for the whole body in total. The underside of the body (throat, breast, belly) except the area around the vent is not included, because feather loss in this region mostly has no or minimal welfare relevance.

Score 1: No wear, good feathering;
Score 2: Moderate wear, i.e. damaged feathers (worn, deformed) or one or more featherless areas < 5 cm in diameter at the largest extent;
Score 3: At least one featherless area > 5 cm in diameter at the largest extent.

Figure 2.7 presents an illustration of damaged feather.

While it is possible to calculate an average plumage score per farm (with a higher score being better), we propose for reasons of transparency to give the percentages of birds with score 2 and 3.

We furthermore propose two alternative sampling strategies: a) Scoring from a distance of up to 2 m. b) Inspection of individually caught hens. The decision for one of the strategies should be based on the selection of further clinical measures.

2.4.3 Body Cleanliness

Validity

Maintaining a clean plumage is not only important for the bird's thermoregulation, physical protection and possibly individual recognition (for birds in small group), but also because the feathers must be in good condition to allow the birds to fly effectively. This importance is reflected in the relatively high proportion of the day spent preening and dust-bathing; for laying hens this has been found to be between 10 and 18 % of the day in one study (Fölsch, 1981), and preening alone as well over 10 % in another (Dawkins, 1989). However, dirty plumage is of greater significance for broilers than for laying hens. Although non-cage housing conditions may lead to dirtier plumage compared to cages, the degree of soiling does not seem to be high enough to impair the functions mentioned above. Consequently in the literature only very rarely plumage cleanliness and feet cleanliness has been assessed in laying hens (Gunnarsson et al., 2000; Wall et al., 2002).

FIGURE 2.7 Damaged feathers.
Source: M. Staack, University of Kassel (photo).

Proposal

Because of the low validity of this measure for laying hens at the moment, we do not propose to consider it for the welfare assessment scheme. It may be useful in the future to to evaluate this measure for use in specific types of management.

2.4.4 WOUNDS

Validity

Skin injuries, by human analogy, are likely to be a cause of pain and a route for potential local or systemic infection, which will affect welfare adversely. However, there is no published work to associate skin damage directly with welfare. When assessing skin damage it is to some extent possible to differentiate causes such as feather pecking or aggressive pecking by the location of the damage. Pecking damage on the comb probably reflects that a bird has received aggressive pecks whereas skin damage at tail and wings probably reflects the pulling out of feathers by conspecifics in the course of feather pecking. However, numbers of pecks at the comb are also increased in flocks with cocks, probably from the mating behaviour. Cannibalistic wounds from vent pecking can typically be found around the cloaca (Savory, 1995) and wounds from toe pecking at the toes.

Assessment Schemes Used in Different Studies

Recording of injuries is typically done concurrently with feather scoring. Sometimes feather and skin damage is even scored with the same rating system, with wounds receiving the highest scores (e.g. Hughes and Duncan, 1972; Allen and Perry, 1975; Bestman and Wagenaar, 2003; Bright et al., 2006). However, the occurrence of wounds is not necessarily linked to feather damage. Gunnarsson et al. (2000) recorded the presence or absence of

scratches, pecks and other lesions for each of 11 body parts as applicable (and without allocating scores). Bilcik and Keeling (1999) defined a rating scale from 0 to 4 with 0 representing no injuries or scratches, and limits between scores being 5 pecks or scratches and regarding wounds 1 cm and 2 cm diameter. A wound of less than 1 cm in diameter received the same score as more than 5 pecks or scratches. Abrahamsson et al. (1996a) assessed wounds only at the comb (pecks and scratches) and around tail/cloaca (score from 1 to 4, 4 being best). Tauson et al. (2005) propose to record pecking damage only at the rear body and comb with a four-point scale which is illustrated by photographs.

Reliability

Gunnarsson et al. (2000) in their comparisons between three assessors found mostly excellent agreements (PABAK>0.8), sometimes for one assessor pair only good agreement (PABAK>0.6 for legs, back between wings, comb) or moderate agreement (PABAK>0.4 for comb). In one case no acceptable agreement was reached (rump part of back PABAK=0.29).

Feasibility

For concerns see feasibility of plumage condition in laying hens. In order to score wounds reliably, birds have to be picked up, and feathers need to be pushed back. Pecking damage, moreover, cannot reliably be seen from a distance of up to 2 m. Even if only wounds of greater diameter are recorded they may not be visible in standing birds when they are at the cloaca or covered by wings/feathers. Assessment of wounds at the slaughterhouse is no alternative, as no differentiation between wounds inflicted on the farm and during catching, transport and waiting at the slaughterhouse can be made. We know, however, from own unpublished data that usually there is a significant correlation between plumage condition and number of wounds, although birds with good plumage may have severe wounds and vice versa (see Figure 2.8). In agreement with this Kjaer and Sørensen (2002) report that feather score and deaths from cannibalism are correlated.

Proposal

Within the framework of the EU-Laywel project a proposal was made to score pecking damage to skin of rear body and comb on a four-point scale (Tauson et al., 2005). We agree with this except that for welfare assessment we propose to simplify the scores for the rear body (including legs) to:

Score 1: No wounds >1 cm in diameter and <3 pecks (punctiform damage <1 cm diameter, see Figure 2.9) or scratches;
Score 2: At least one wound > 1 cm diameter or ≥ 3 pecks or scratches;
Score 3: At least one wound > 2 cm diameter.

FIGURE 2.8 Hen with moderate wear for plumage condition (score 2), but severe wound.
Source: M. Staack, University of Kassel (photo).

Notice that in the EU-Laywel project the scoring system proposed is the reverse of this one, i.e. a low score means high damage and a high score means low damage. The reason for 'reversing this' is to bring it into agreement with the overall trend in Welfare Quality® which is toward low scores for 'best welfare' conditions.

While it is possible to calculate an average skin score per farm (higher is better), we propose for reasons of transparency to give the percentages of birds with score 2 and 3. For the comb only the percentage of birds with equal to or more than 3 pecks should be recorded. These proposals apply to individually captured birds. For standing birds only score 3 can be recorded for rear body.

Wounds could also be measured on slaughter line. In this case just presence or absence of wounds > 2 cm (rear body) and pecks at the comb should be recorded.

2.4.5 FEET CONDITION

Validity

Alterations of foot pads in laying hens can occur as hyperkeratosis or ulcerative dermatitis (bumble foot). Factors favouring especially bumble foot are among others wet litter, inadequate perches, skin damage, such as cuts, abrasions or erosions and the presence of pathogens (e.g. Staphylococcus), but also genotype such as white hybrids (Tauson and Abrahamsson, 1996; Wang et al., 1998; Wahlström et al., 2001; Vits et al., 2005). During the inflammatory process, fluid may be produced around the joints and tendon sheaths. Wounds heal on the outside to leave a hard core of pus in the inside. If the condition becomes chronic, fibrous tissue can form around the foot. It is very likely that these processes are painful to the birds. They also increase the risk for toe cannibalism. The welfare impact of hyperkeratosis is less clear.

Figure 2.9 Wounds caused by pecks at vent/belly.
Source: M. Staack, University of Kassel (photo).

Assessment Schemes Used in Different Studies

Also for the assessment of feet condition in live hens on farm the scoring systems used differ to some extent. Abrahamsson et al. (1996a, b) used a four-point scale according to the presence of hyperkeratosis in the distal toe pad and abscesses, but also of 'bumblefoot' in the foot pad. Weber et al. (2003) used a 5-point scale taking into account proliferation, degree of necrosis and swelling. Gunnarsson et al. (2000) distinguished at the foot pad: no lesion, crust, excessive horn layer, papillae, bleeding chaps, acute bumble foot, chronic bumble foot and corn-like lesions. Oester (1994) differentiated three degrees of bumble foot, two degrees of a preliminary stage and no alterations.

Reliability

Gunnarsson et al. (2000) reached excellent to good agreements between two pairs of assessors (PABAKs=0.94, 0.75). Whereas in meat birds dirtiness of the feet mostly impairs reliability considerably, in laying hens this problem exists to a much lesser degree.

Feasibility

The same concerns apply as expressed in feasibility of plumage condition in laying hens. Feasibility of the foot assessment is better at the slaughterhouse post mortem than on the farm. However if the decision is taken to individually capture and inspect birds, the extra time needed for the assessment is minimal and, therefore, the scoring should be carried out then as well.

Proposal

Within the framework of the EU-Laywel project a proposal was made to score 'bumble foot lesions' on a four-point scale (Tauson et al., 2005). For the welfare assessment we propose to simplify the scores to a three-point scale:

Score 1: Feet intact, no or minimal proliferation of epithelium;
Score 2: Necrosis or proliferation of epithelium or chronic bumble foot with no or minimal swelling;
Score 3: Swollen.

As for the scoring system for plumage condition, notice that in the EU-Laywel project the scoring system proposed was the reverse, i.e. a low score means high damage and a high score means low damage.

Foot lesions could also be measured on slaughter line with the same scoring system as described above.

2.4.6 Keel Bone Condition

Validity

Keel bone deformations are due to osteoporosis in combination with collisions with housing fittings and/or long-term pressure (Vits et al., 2005) leading to fractures or bone deformations. 73% of birds in a perchery system had old breaks at the end of lay (Freire et al., 2003) and in another, between 50% and 78% of birds in free range and barn housing systems were found to have healed fractures of the keel and furculum (Wilkins et al., 2004). Chickens have been shown to have sensitive pain perception mechanisms (Gentle 2001) and so are likely to experience pain from bone fracture, which will also reduce bird mobility. The welfare impact of mere bone deformations is less clear. However, Scholz et al. (2008) showed on the basis of histological examinations that 100% of hens with a high degree of keel bone deviations (assessed visually and by palpation in live birds) had fractures, 80% of hens with medium deviations and still 50% of hens with slight deviations.

Assessment Schemes Used in Different Studies

Keel bone deformation has been assessed in live laying hens for example by Abrahamsson et al. (1996b) or Wahlström et al. (2001) using scores from 1 to 4 (with 4 is best), but they gave no explanation on the definition of these scores. Gunnarsson et al. (2000) differentiated between straight keel bones and those deviating less than 0.5 cm, 0.5-1 cm and more than 1 cm.

Reliability

The predictive value of the palpation technique, that is the probability of a bird having a broken bone if one was diagnosed by palpation, was found to vary from 74% to 88% depending on the observer and all observers recorded over 70% correct assessments (Wilkins et al., 2004). Gunnarsson et al. (2000) found excellent to good agreement between two pairs of assessors for the assessment scheme (PABAKs=0.88, 0.73).

Feasibility

Visible examination of bones gives a reliable assessment measure of the number and severity of fractures sustained. However, this method is not feasible routinely as it is an extremely time consuming procedure and birds have to be killed. For individual palpation in live birds the same concerns apply as expressed in feasibility of plumage condition in laying hens. When keel bones are assessed by palpation at the slaughterhouse, no differentiation between damage originating from the farm and from catching, transporting and handling at the slaughterhouse can be made. Dissection of slaughtered birds would allow differentiation between old and new fractures, but this will be too time-consuming. If the decision is taken to individually capture and inspect birds, the extra time needed for the keel bone assessment is minimal and, therefore, the scoring should be carried out.

Proposal

We propose that the number of birds with keel bone deviations is recorded if the birds are individually picked up and assessed.

2.5 DISCUSSION

Taking reliability and feasibility aspects as well as the possible meaning of the results for the welfare assessment issues into account, we generally propose a simple scoring scale with 3 or a maximum of 4 points or a present/absent assessment. The lower the differentiation between scores the better the reliability that can be expected. Not to have to differentiate between body regions increases the speed of assessment. If the results of the plumage condition and wounds assessment in laying hens are to be used for advisory purposes, differences between body regions may be of importance. However, for a general welfare assessment this is not the case.

For the time of the assessment it should be considered that in broilers the end of the rearing period (2–3 last days) could be the best time window. It is probably at that time that body

cleanliness, breast blisters, but also hock burns are the worse. However, the assessors have to be aware that for pododermatitis some healing seems to appear during the rearing period and, then, it is not necessary at the end of the rearing period that the severity of the pododermititis is the higher (Gordon et al., 1996). In laying hens the peak of cannibalistic activity is around the peak of lay whereas plumage damage as well as keel bone alterations consistently increase until the end of the laying period (Huber-Eicher and Sebö, 2001; Wahlström et al., 2001). For foot alterations conflicting patterns are reported (Abrahamsson et al., 1996b; Wahlström et al., 2001). As already in week 32 a considerable increase in feather damage was repeatedly noted (Bilcik and Keeling, 1999; Huber-Eicher and Sebö, 2001), we propose that the best time window to assess the laying flocks is around week 32 to 38.

3

STANDARDISATION OF MEASURES OF BROILER LAMENESS

A. Butterworth and C. Arnould

3.1 SUMMARY

Leg health problems are to be found globally in the production of modern broilers. Studies in Denmark and Sweden have found high levels of leg health problems in broiler chicken flocks (Sanotra et al 2003). These authors found levels of birds with gait score of over 3 to be 30.1% in Denmark, for Ross 208 birds, and 14.1% and 26.1% in Ross 208 and Cobb birds, respectively, in Sweden. The findings of the 2005 UK study (Anon, 2006), as well as several others, indicate that that one of the major causes of leg health problems in broiler chickens is high body weight gains due to fast growth rate (Kestin et al 2001; Sanotra et al 2003), which may be genetically determined (Kestin et al 1992; Kestin et al 1998; Bradshaw et al 2002; Julian 2005). A short outline of available methods for on farm assessment of lameness is provided. This report identifies the process used to make a decision on a (current) pragmatic methodology for assessment of lameness on farm, and provides the basis for a methodology for assessment on farm which will be tested and refined in the ongoing work packages in sub project 2.2.

3.2 INTRODUCTION

Lameness in meat birds is recognised as a serious welfare problem. As the problem is associated with rapid growth rate at an early stage in the birds development (Kestin et al., 2001), lameness is of minimal concern for laying hens. Levels of leg disorders which lead to lameness can be high in rapidly growing broilers (Sanotra., 2000). This has been highlighted as a major welfare concern (FAWC 1992; FAWC 1998; European Commission 2000). However, there has been debate about the level of leg disorders in UK, European and world flocks (<http://www.fawc.org.uk/letters/broilerlet.htm>). Leg weakness can result from metabolic, developmental (Williams et al., 2000) or infectious causes

(Butterworth, 1999). In addition, developmental abnormalities can act as a focus for infection (McNamee et al 1998) and some developmental abnormalities are exacerbated by nutritional imbalances such as choline deficiency (Ferguson et al., 1978) and the effects of including soya bean oil in the diet (Watkins et al., 1991) Genetic factors are also an important cause of leg weakness (Le Bihan et al., 1996; Kestin et al, 1992).

3.2.1 Potential Assessment Methods for Lameness

A recent study by Bizeray et al. (2000) examined locomotor behaviour in different genotypes of chicken with different growth rates, and used time budget measures with a focal and scan sampling techniques.

Reiter and Bessei (1997) used video tracking of reflective points on marked birds to compare vertical and horizontal movements in sound (no lameness) and lame meat chickens and in layers. However, no further categorisation was attempted, the method being described as a tool contributing to quantitative description of gait abnormalities.

Kestin et al. (1992) described a subjective scoring system based on a scale of 0 to 5, with precise description of the gait and walking criteria to be met by birds within scores. This scoring system is known as the Bristol Gait Scoring System (BGSS). Six categories are defined, from Gait Score 0 (zero), 'the bird walked normally' to Gait Score 5'the bird is incapable of sustained walking'. Gait scoring is usually carried out by two people, one herding the birds with a light stick, and the other recording observations.

Recent American studies have refined the existing gait scoring method (BGSS), to give a Modified Gait Score System (MGSS) (Garner et al., 2002). The MGSS add objective inclusion/exclusion criteria for each score compared to the BGSS. Dawkins and co-workers used a three point lameness scale in recently published work on the effects of stocking density on broiler welfare (Dawkins et al., 2004). Weeks et al (2002) described an objective method for assessing lameness based on the length of time that birds remain standing in shallow water. This Latency to Lie (LTL) test was modified by Berg and Sanotra (2001) to enable faster testing of individual broilers.

Locomotor analysis can be carried out through the recording of ground reaction forces measured from a force plate. A force plate is a platform where ground reactions are measured with strain gauges when an individual walks on it. This method has been used to detect lameness in broilers previously classified as either lame (Gait score = 1-3) or not (Gait score = 0) using the gait scoring system (Sandilands et al., 2003). Two descriptors of walking style correctly identified 90% of the birds as either lame or not lame. However the automation of the kinetic analysis is very difficult and the use of a force platform is therefore limited to experimental studies.

Invasive techniques which measure bone and joint deformity post mortem (Riddell, 1976; Riddell and Classen, 1992) are not feasible for welfare assessment purposes due to complexity and time constraints.

3.2.2 Validity and Repeatability

It is intuitive that birds which have difficulty walking are likely to be experiencing pain and/or discomfort. Lameness, levels of which are increased by fast growth rate, has been reported to reduce levels of: visits to feeders and time spent standing and eating (Pattison, 1992; Weeks and Kestin, 1997; Weeks et al., 2000); ground pecking (Weeks and Kestin, 1997); preening (Weeks et al., 2000); and, standing and scratching (Reiter and Kutritz 2001). An increase in the number of birds showing abnormal lying posture (Weeks and Kestin, 1997), reduced running and walking (Weeks and Kestin, 1997; Weeks et al., 2000; Reiter and Kutritz, 2001) and reduced levels of dust bathing (Vestergaard et al., 1997; Vestergaard and Sanotra 1999) have also been reported in birds with leg weakness.

Of the methods for assessing walking ability, validity, repeatability and feasibility have only been evaluated for the BGSS, the MGSS and the LTL. Other methods described may be of value in a research context but have not been assessed in the context of the assessment of bird walking ability on farm. Consequently, only the BGSS, LTL and MGSS will be examined further here.

Studies have shown that birds with high gait scores negotiate an obstacle course more rapidly after Non-Steroidal Anti Inflammatory Drug administration (McGeown et al., 1999), and preference testing reveals that they have a greater tendency than non-lame birds to select feed treated with analgesics (Danbury et al., 2000). These studies demonstrate that birds with BGSS scores of greater than 2 may be experiencing pain as well as impaired mobility. Leg weakness may be due to skeletal deformity or due to joint or bone infections (Butterworth, 2000).

The test-retest reliability of the BGSS has been found to be 0.7 in one study (Kestin et al., 1992) and 0.906 (Garner et al., 2002) in another; that for the MGSS has been found to be 0.948, for naive observers (Garner et al., 2002). Berg et al. (2005) have more recently found inter-rater reliability for the BGSS between researchers and between industry raters to be good but agreement between the two groups was poor. Garner et al. (2002) found that inter-rater reliability was 0.892 for the BGSS and 0.943 for the MGSS. Thus, reliability of these two gait-scoring systems is good, and, from this limited published work, the MGSS seems to show slightly better test-retest and inter-rater reliability.

3.2.3 Feasibility

In an on farm assessment context, the BGSS has been found to provide a feasible technique for the assessment of walking ability. A large numbers of 'scores' may be obtained in a comparatively short time, which increases confidence in the results obtained for a specific flock. Garner et al. (2002) found that there was no significant difference in the number of attempts made to score each bird or the difficulty of scoring them between the BGSS and the MGSS. Any visual gait-scoring scheme performed by an observer involves a degree of subjectivity in the scoring, in common with any numerical rating system of a continuous variable. However, both the BGSS and MGSS are feasible for use in an on-farm audit

because they are non-invasive methods, used on live birds, and have minimal influence on 'normal' walking behaviour, although the MGSS as described by Garner et al. (2002), involves picking up birds prior to assessment of walking ability, which has the potential to alter bird behaviour, depending on the perceived stressfulness of this procedure to the bird . No complex machinery or imaging system is required for either the BGSS or the MGSS and pre -examination preparation is minimal.

The BGSS has been used experimentally to determine factors associated with high levels of bird lameness and thus enable industry to adopt husbandry systems and genetic selection methods that reduce lameness levels (Sorensen and Kestin, 1999; Su et al., 1999). BGSS is used in farm accreditation audits in the UK, for example (RSPCA, 2000), and also as an aid to guide on-farm culling policy for lame birds.

Comparisons of the LTL test with the BGSS show clear negative correlation with LTL ($P<0.001$) in both reported studies of broilers tested on commercial farms (Weeks et al 2002, Berg and Sanotra, 2003). The latter study also found the mean LTL values for the different gait scores to be significantly ($P<0.01$) different and no significant difference in LTL results between 14 flocks. Berg and Sanotra (2003) modified the LTL test for use on commercial farms and concluded that it was well suited for on farm use, thus demonstrating feasibility. Neither paper reported on the inter-observer reliability for LTL, but this test is inherently reliable and repeatable as, in contrast to BGSS and MGSS, it is a purely objective test, requiring no subjective interpretation by observers.

3.2.4 The Evaluation of the Potential Methods

After discussion and a review of available information, a list of potential lameness assessment methods was compiled (Table 3.1). Each method was ranked using the following method. A short line (VAL visual analogue line) was created and the score (0 to 5) calculated from the position marked by the partner on the line. Four lines were used; they corresponded to the questions 'How critical overall is the issue for animal welfare', 'The time needed to get the data', 'Simplicity of measure', and 'Quality/reliability of data'. The total score was calculated for each group of questions and the average for the partners calculated (minimum score: 4, maximum score: 20). This comparatively uncomplicated method allowed a summary view of the relative merits and disadvantages of each possible method to be created. It is accepted that this method only provides a very simple 'pointer' for the practical merits of the methods described. The members of the group were chosen for their experience in this area so there had already been a large amount of filtering of questions.

3.2.5 Decision on Method to Propose

On the basis of this decision making process, it was decided that visual gait scoring using a set of recognised gait score categories would provide the most appropriate mixture of time, technical and specificity as well as being considered practical for on farm assessment

TABLE 3.1 Decision tool for potential lameness assessment methods.

Criteria	Advantage	Disadvantage	Overall averaged score
Time budgets	Indicate the direct effect on the individual animal, allows precision regarding specific effects on behaviour and the complex interactions between these.	Time consuming. Potentially technically demanding (video analysis, etc.)	10
Video tracking	Automation, precision (?) repeatability, reliability, reduced inter observer variability (?)	At present only limited experimental use. Not yet practically available for on farm use.	14
Force plate studies	Automation, precision (?) repeatability, reliability, reduced inter observer variability (?)	At present only limited experimental use and birds need to be trained / forced to walk over the force plate.	12
Latency to lie.	Can give group results without the need to make individuals move /walk. May represent a direct link to difficulties / aversion of standing. Of high value to validate other 'large group' methods.	May be time consuming to test reasonable numbers of animals. Technical / ethical resistance to the use of warm / tepid water on birds in a commercial setting.	16
Gait scoring (primary method = Kestin et al 1992) (modified method = Garner et al 2002)	Can be realistically used on reasonably large sample sizes. Trained observer can give reasonable repeatability / reliability. Already widely adopted.	Potential for inter observer variability / bias. Random selection of birds important. Score range (0 to 5) can offer precision but also can lead to difficulties in categorisation, particularly in the low score areas 1,2,3,	18.5
Gait scoring (Dawkins 2004) (Modified gait scoring scheme)	Reduced scoring categories (3 scores) simplified the observation and decision making.	Reduced scoring categories may result in reduced precision.	17.5

of lameness. This is in line with the commercial finding that poultry producers have been able to assess lameness within their own companies using personnel trained in gait scoring techniques. No commercial company has adopted automated or video gait scoring techniques, but a number, including companies in Asia and South America have received training in gait scoring as a tool to assist in the management decisions on animal welfare and disease within their own companies.

3.3 DISCUSSION

To allow some 'measure' of the severity of lameness, and hence to allow improved understanding of the effects of lameness on a flock basis, a number of studies have given subjective severity 'scores' to lameness in commercial birds.

Reiter and Bessei (1997) used videotracking of reflective points on marked birds to compare vertical and horizontal movements in sound and lame broilers and in layers. However, no further categorisation was attempted, the method being described as a tool contributing to quantitative description of gait abnormalities.

Kestin et al. (1992) described a subjective scoring system based on a scale of 0 to 5, with precise description of the gait and walking criteria to be met by birds within scores. Gait scoring, as described, is usually carried out by two people, one herding the birds with a light stick, and the other recording observations. Kestin et al. (1994) describe the use of this method for gait analysis of commercial birds, and obtained a high repeatability for scoring of the same birds twice on the same day. Recent American studies (Mench, 2002 and pers comm.) have started to refine the existing gait scoring method.

Whilst accepting the subjective nature of any gait scoring scheme, the advantages of such a system are:

1) Non-invasive methods, such as gait scoring, have the least influence on 'normal' walking behaviour of the birds.
2) The birds may be assessed alive. Some previous categorisations have relied on measurement of bone and joint angles after dissection (Riddell, 1976; Riddell & Classen, 1992).
3) Large numbers of birds may be gait scored in situ, i.e. in the poultry house.

If practical welfare steps based on lameness are to be considered on farm, a critical level may be set by the stockman i.e. gait score 4 & 5, and birds reaching this level destroyed humanely on farm.

4) No complex machinery or imaging system is required.

3.3.1 Protocol for Gait Scoring on Farm

A recognised method for assessing the walking ability of the birds was developed by Kestin et al (1992) and this forms the basis for the assessments. Birds are given a score from 0 to 5, with 5 being the poorest gait. The following written descriptions of Gait score can be taken as a guide to the broad description of the behaviour of animals in the different gait score categories – however – training using video footage, and on farm assessment of

the performance of the assessor is considered to be more robust, giving assurance as to the uniformity, repeatability and validity of the assessor performance.

For gait scoring, birds should be selected at random within the house by reference to a randomised location identifier. Birds are selected from 4 locations, in groups of up to 80 birds, by corralling at each location using a hinged catching pen. Each bird is individually encouraged to walk out of the pen and is scored as it does so. Birds should not be lifted or moved manually as this may disrupt their normal walking behaviour.

3.3.2 Calculation of Flock Average Gait Score

Flock average gait score is a 'single reference measure' which can be used to compare lameness incidence rates between individual houses, flocks, farms and companies. An example of how it is calculated can be found in Table 3.2.

TABLE 3.2 Example of calculation of Flock Average Gait Score (FAGS). FAGS can be used as a single measure to compare lameness on individual farms, flocks and within companies.

A Gait score category	B Number of birds scored in this category	C % of total	D = A x B
0	10	4	0
1	25	10	25
2	137	55	274
3	60	25	180
4	11	4	44
5	5	2	25
	Total of B scored = 248	Total = 100 %	Total of D = 548
		Flock average = Total of D / Total of B scored = 548 / 248 = 2.21	

4

MEASURES OF POULTRY HEALTH STATUS

A. Butterworth and K. Niebuhr

4.1 SUMMARY

Disease can be an unfortunate but inevitable occurrence in farm animals. Good health, appropriate freedom from disease and injury are basic requirements regarding animal health and welfare. Welfare assessment protocols designed for the assessment of individual farms may take into account parameters which directly or indirectly reflect the health of the birds. Data regarding health are sometimes recorded on a routine basis, yet may not be readily available, or may be recorded in variable formats. Thirty four potential health status measures (some of which contain subconditions) are considered a decision support tool is used to rank the measures to provide some ability to discard less highly ranked measures if required. Simple measure descriptions are given and where appropriate scoring schemes are proposed.

4.2 INTRODUCTION

Disease can be an unfortunate but inevitable occurrence in farm animals. Good health, appropriate freedom from disease and injury are basic requirements regarding animal health and welfare. Good health, freedom from disease and injury and ultimately survival in a given flock are basic requirements regarding animal welfare. With regard to poultry, health status has been widely discussed in recent years, especially problems with the potential for high mortality, such as that due to infectious diseases or vent-pecking in laying hens, injuries such as feather- and vent-pecking, bone fractures and bumble foot (Huchzermeyer et al., 1986; EU, 1997; DECCA, 2001; Hafez et al., 2001). Therefore any welfare assessment protocol designed for the assessment of individual holdings must necessarily take into account parameters which directly or indirectly reflect the health and survival of the birds. In most cases these problems are not influenced solely by housing

criteria but are multi-factorial (Huchzermeyer et al., 1986; EU, 1997; Hafez et al., 2001), including inter alia management in general, feeding, rearing, and, preventive and curative interventions. Thus, ideally, parameters which reflect health status, including mortality rates, might be included in a flock welfare assessment protocol. Although the validity of health parameters regarding animal welfare is recognised, few studies regarding the inclusion of health parameters in welfare assessment protocols, except for integument scoring for studies on feather- and vent-pecking, have been published. Disease prevalence may be assessed on-farm through examination of live birds, post mortem examination of either culled or dead birds (at the slaughter plant), laboratory examination of samples from live or dead birds and study of disease and medication records (if available). Live bird assessment might include, palpation of the sternum, to assess deviations and fractures (Gunnarsson et al., 1995), and examination for signs of diarrhoea or nasal discharge. Post mortem examination findings of culled or dead birds, on the day of inspection, might be also included in a welfare assessment protocol (Hegelund et al., 2003).

Histological, microbiological and parasitological investigations of live and dead birds are routinely conducted in response to disease outbreaks. The results of these tests may provide information about the severity of the disease condition in terms of bird welfare. Data regarding health are already often recorded on a routine basis, yet may not be readily available. This is true for data compiled by governmental animal health services (at least in some countries) and by local veterinarians, whose data may include data on faecal examination for parasites, preventive interventions like vaccinations, treatment and prescription of medication, or results of post mortem examination. Clinical examination, clinical post mortem, laboratory analysis of samples and the maintenance of health records are routine procedures on poultry units, conducted largely by veterinary specialists. Consequently, capturing such data for welfare assessment protocols may be feasible in integrated poultry businesses. However, currently, health records may not be readily available for assessment purposes and may be held on databases remote from the farm.

A shortcoming of any more detailed examination and evaluation of health data as currently gathered is that it may require a trained specialist to interpret the results. It can be safely assumed that, given institutional safeguards, the data generated by laboratories and specialists can be structured to demonstrate a high order of reliability. However, currently, procedures for recording such data are not standardised and therefore not directly comparable between production units and companies; reliability might therefore be enhanced by standardisation.

4.3 METHODS AND RESULTS

4.3.1 Potential Assessment Methods for Health Status

Initial discussion focused on the need for a tabulation of possible health status measures for poultry to ease the understanding of the variety of possible measures. Table 4.1 was used to help understand the strengths and weaknesses of the possible health status measures, the features of each measure were listed, where they could be applied (Farm or slaughterhouse) and then visual analogue lines used to score the merits of each potential measure using the criteria below (A to D). Although other criteria could have been selected, these (below) appeared to give a simple method for ranking the measures. The partners in this subproject used the table format below to help in their decision making. Higher scores indicate a better mix of welfare significance, practicality, data collection ease and specificity and the significance of critical windows for assessing the specific measure.

Each method was ranked using the following method. A short line (VAL visual analogue line) was created and the score (0 to 5) calculated from the position marked by the partner on the line. The total score was calculated for each group of A to D (Table 4.2) and the average for the partners calculated. This comparatively uncomplicated method allowed a summary view of the relative merits and disadvantages of each possible method to be created. It is accepted that this method only provides a very simple 'pointer' for the practical merits of the methods described, and it will be for ongoing studies in WP2.2 (on farm application and assessment of feasibility) to determine the practicality of the measures within a monitoring scheme. The members of the group were chosen for their experience in the area of on farm assessment of poultry.

4.3.2 Discussion on the Measures

The partners had significant discussion about the role of the measures – were they to be used predominately in the allocation of farm integrated scores for the labelling and certification process, or were they also to be used for the feedback to the farmer in an advisory /supporting role? The discussion focused on the different attitude that the farmer might take to measures used in each category – for example, if disease incidence was given a negative score – then farmers might consider;

- That disease should not be fully recorded or reported
- That not only did disease cause welfare and economic problems but also manpower and resource demands

There are measures that may influence farmer compliance – for example, a discussion on the problems of mortality recording. Overall, a reduction in mortality is considered an appropriate goal. However, mortality is comprised of controlled and uncontrolled elements,

TABLE 4.1 Decision tool for potential health status assessment methods. Visual analogue line (VAL) (scored 0 to 5) used to provide a basic ranking of measures. (VAL marks not shown as they differ for each partner – average score shown.

Criteria/clinical measure	Notes Best suit certification (C)/advisory (A)	Layer/broiler?	Where to assess?	Links to other measures in WQ
Enlarged crop	C (B) A (L)	B/L	F (S if birds with enlarged crops not removed to avoid contamination)	
Dehydr.	C	B	S	Management measures and Stockmanship
Emaciation	C Not very useful for white layers	L/B	S	
Pericarditis	C	L/B	S	
Ascitis	C	B (L?)	S (sudden death on farm)	
Deep pectoral myopathy	C, A	B	S	
Hepatitis, Jaundice	C Fatty liver in laying hens	L/B	S (colour of birds on-farm)	
Septicaemia	C	L/B	S (PM) F (sick & moribund birds?)	
Lameness assessment (gait scoring)	C, A	B	F (gait scoring) S (pathologies)	Lameness Management measures (litter management, culling) and Stockmanship
Synovitis, Tenosysnovitis	C, A	B	S	Lameness Management measures (litter management, culling) and Stockmanship
Hock burn	C, A	B	S	Feet condition Lameness Management measures (litter management, culling) and Stockmanship
Pododermatitis	C, A	B	S / F	Feet condition Lameness Management measures (litter management, culling) and Stockmanship
Spinal cord pathologies	C, A (likely to be rare)	B	F (behaviour)	Lameness
Bone pathologies – angular bone deformities	C, A	B	F (angular deforities) S (dissection)	Lameness

Notes: A x _____ xxx How critical is the issue for animal welfare (low to high)
B xxx _____ x Time needed together the data (quick or time consuming)
C x _____ xxx Simplicity of the measure)? (less – more)
D x _____ xxx The quality/reliability of the data (poor to good)

TABLE 4.1 CONT. Decision tool for potential health status assessment methods. Visual analogue line (VAL) (scored 0 to 5) used to provide a basic ranking of measures. (VAL marks not shown as they differ for each partner – average score shown).

Criteria/clinical measure	Notes Best suit certification (C)/advisory (A)	Layer/broiler?	Where to assess?	Links to other measures in WQ
Salpingitis (inflammations of the reprod. tract)	C	L	S (viscera on line)	
Enteritis	C, A May be readily measured in broilers, but less easily in brown layers	L/B	F (diarrhoea, litter quality, bird cleanliness)	
Skin damage (pecking damage)	C, A Critical window (layers) between weeks 23 and 45	L (B? quite rare in broilers)	F (critical window week 23 to week 45)	Wounds
Abrasive feather damage	C, A	L	F	Plumage
Feather damage caused by pecking	C, A	L	F / S	Plumage
Cleanliness	C (B) A (L)	L/B	F (if able to pick up birds – broilers) S (if not able to pick up)	Plumage Cleanliness
Foot damage (injuries caused by environm.,	C	L	S (low prevalence?)	WoundsFeet condition Lameness Management measures (litter management, culling) and Stockmanship
Toe damage (caused by injurious pecking)	White layers only C	L	F, S	Wounds Feet condition
Breast blisters	C, A	L/B	S F (if catch birds)	Breast blisters Management measures (litter management, culling) and Stockmanship
Wing fracture	C	L (B?)	S	
Limb fracture	C	L (B?)	S	Lameness
Keel bone fracture	C	L	S (if after depopulation + period of lay) F (for period of lay only)	
Comb wounds	C	L	F	
Eye pathologies (Keratoconjunctivitis, opaque cornea)	C, A	L/B	F	

Notes (cont.): B = Broiler chicken, L= Layer, C = certification (farm inspection), A = advisory/supporting role (for far-mer advice or information), F = Farm, S = slaughterhouse

TABLE 4.1 CONT. Decision tool for potential health status assessment methods. Visual analogue line (VAL) (scored 0 to 5) used to provide a basic ranking of measures. (VAL marks not shown as they differ for each partner – average score shown).

Criteria/clinical measure	Notes Best suit certification (C)/advisory (A)	Layer/broiler?	Where to assess?	Links to other measures in WQ
Nasal discharge upper respiratory	C, A	L/B	F	
Dermatitis, cellulites (not pododermatitis)	C	L/B	S	Management measures (litter management, culling) and Stockmanship
Parasites – red mite infestation	C	L	F	
Endo-parasites (Faecal egg / coccidian counts may indicate disease challenge)	C, A	L/B	F (Laboratory)	
Records	Include lab. test results for disease monitoring Vaccination Antibiotics Vitamins / miner-als	L/B	F	Management measures (litter management, culling) and Stockmanship
Mortality	See discussion on mortality below**	L/B	F,S (for cross checking of mortality information)	Dead on arrival

and a more focused goal would also be to maximise the element of mortality that is controlled (i.e. managed humanely by the producer, not left to 'chance').

If farmers perceive that mortality is being used as a punitive measure for their performance then they may find ways of providing modified information, for example by altering the ratio of culls to found dead birds. For an advisory role mortality should be split into cull (controlled mortality) versus found dead (uncontrolled mortality). For the coarsest measure, mortality can be set at an overall threshold – for example, 8% over total flock life – but this measure will tell you nothing about the welfare of the animals that died, simply that they died. But did they die in a controlled way (humanely culled) or did they die undetected in a cage / in the litter without intervention from the stockman?

If it is decided that a measure is best (most easily) recorded at the slaughter plant, then there may be a need for care to avoid farmers from always removing birds showing low level lesions before sending the birds to slaughter to avoid any score for these low level conditions (especially for smaller flock sizes). An example would be birds with low levels of hock burn, or with skin lesions such as cellulitis.

TABLE 4.2 Rank order of measures. The respondents average score for each of the clinical measures described in Table 1, using the scoring criteria A,B,C& D described in Table 1 is ranked in decreasing order.

Clinical measure	Score	Rank
Ascites	23	1
Hepatitis, Jaundice	22	2
Hock burn	21.5	3
Keel bone fracture	21.5	3
Pericarditis	21	4
Salpingitis	21	4
Feather damage caused by pecking	21	4
Comb wounds	21	4
Parasites	21	4
Mortality	21	4
Emaciation	20.5	5
Breast blisters	20.5	5
Deep pectoral myopathy	20	6
Records	20	6
Septicaemia	19.5	7
Pododermatitis	19.5	7
Foot damage	19.5	7
Wing fracture	19.5	7
Cleanliness	19	8
Toe damage	19	8
Dermatitis, cellulites	19	8
Dehydration	18.5	9
Enteritis	18.5	9
Limb fracture	18.5	9
Skin damage	17	10
Abrasive feather damage	17	10
Eye pathologies	17	10
Endoparasites	17	10
Lameness assessments	16.5	11
Synovitis, Tenosysnovitis	16.5	11
Spinal cord pathologies	16	12
Nasal discharge	16	12
Bone pathologies	14.5	13
Enlarged crop	13.5	14

4.3.3 Assessing the Proposed Measures

Table 4.3 presents the range of proposed measures for assessing the health of poultry at a number of points: slaughterhouse, farm and laboratory.

TABLE 4.3 Assesment of proposed measures in poultry health.

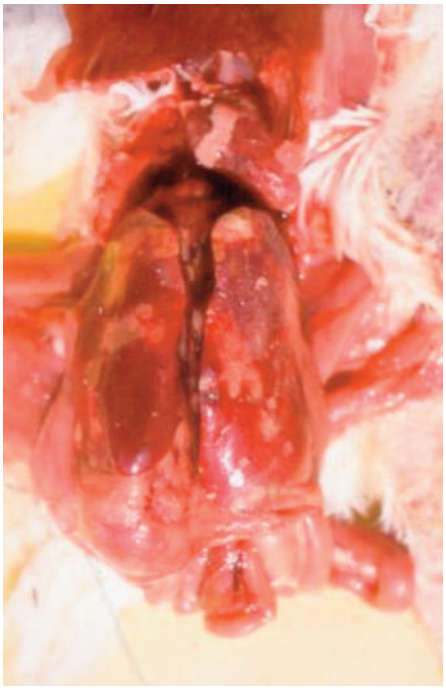
Ascites

S – Examination of viscera and heart – presence of fluid (ascitic fluid, plus dis-tension of abdomen and fluid filled fat and tissues, enlargement of heart / pericardium)

Proposed scoring scheme:
None 0, Mild 1, Moderate 2, Severe 3

Hepatitis, Jaundice

S – Examination of viscera – presence of dicolortion and/or evidence of localised liver damage or infection.
Proposed scoring scheme:
None 0, Mild 1, Moderate 2, Severe 3



Notes: S = slaughterhouse; F = farm; L = laboratory.

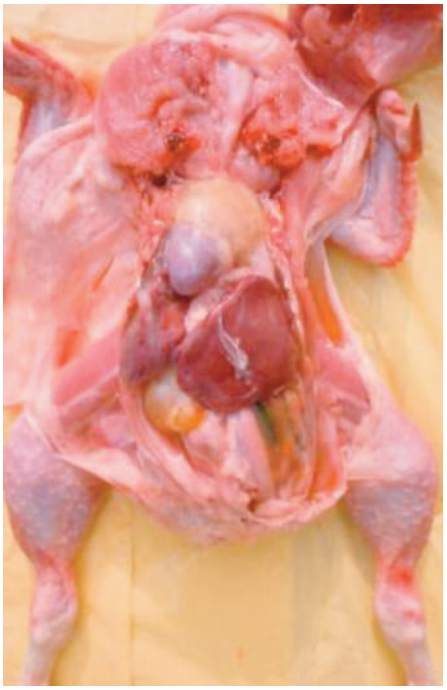
TABLE 4.3 CONT. Assesment of proposed measures in poultry health.

Hock burn

1 2 3 4 5

Gleadthorpe hock burn scale
1: No lesion.
2: Very small and superficial lesions, slight discolouration on a limited area, mild hyperkeratosis.
3: Mild lesion: discolouration of the hock, superficial lesion, superficial dermatitis.
4: Moderately severe lesion: ulcers or scabs
5: Very severe lesion: ulcers or scabs, signs of haemmorrhages or deep dermatitis.

Pericarditis



S – Examination of viscera – presence of distended pericardium with fluid be-tween pericardium and the heart – may be associated with ascites

Proposed scoring scheme:
% none (0)
% mild (1)
% moderate (2)
% severe (3)

Mortality

F = production mortality %, or mortality to a given day (x% at day y)
% mortality – initial stocking should be accurately known, or culls and birds found dead should be recorded separately. Overall % mortality is less important than % of mortality that is not carried out under the control of the stockman – i.e. managed mortality.
S = Dead on arrival (DOA) of those transported arriving dead.

Emaciation

% of flock suffering underweight/undercondition
if emaciation present then:
% none (0)
% mild (1)
% moderate (2)
% severe (3)

TABLE 4.3 CONT. Assesment of proposed measures in poultry health.

Dermatitis, Cellulites

% none (0)
% mild (1) (<2cm dia lesion)
% moderate (2) (2–5 cm lesion)
% severe (3) (>5cm lesion)

Dehydration

Dry tacky carcase at slaughter, some-times discoloured (red) and poor bleedout.
Absent (0)
% Present (1)

Red carcass (dehydrated and poorly bled) Normal

Enteritis

Determined by litter and bird vent soil-ing. May contribute to bird cleanliness score (as above).
Record absence (0) or presence (1) and note if very severe soiling or bird morbidity.
Estimate for % birds affected.

Limb fracture

Absent (0) or present (1) likely to be severe impact on animal if present and to result in culling (B) or bird may sometimes not be recognised and will recover (L) in cage systems.
% of flock affected

Eye pathologies

Unilateral and bilateral ocular pathologies with discharge or discolouration of the cornea
% none (0)
% mild (1)
% moderate (2)
% severe (3)

Lameness assessments

See Chapter 3.

Synovitis, Tenosysnovitis

Usually B but possible L, post mortem (S) to confirm, but possible to palpate birds on farm (F)
None (0)
Mild (1)
Moderate (2)
Severe (3)
And % of flock affeceted.

TABLE 4.3 CONT. Assesment of proposed measures in poultry health.

Spinal cord pathologies

Paralysis with characteristic backwards shuffling walk (B)
% affected

Enlarged crop

(B) % affected

4.4 CONCLUSIONS

Following the work in WP 2.2 and 4.6.3, a refined and developed list of measures was compiled, tested and evaluated on farms in the UK, France, Netherlands, Sweden and Italy. The resulting reduction and refinement led to the following measures being adopted for clinical scoring in broiler chicken taken on farm and at slaughter (Table 4.4) and laying hens taken at farm (Table 4.6). The detail of the measures included in the WQ protocols in June 2009 are given in the linked tables for broiler chicken (Table 4.5) and for laying hens (Table 4.7).

TABLE 4.4 Measures adopted in the 2009 WQ protocols for clinical scoring (poultry health status broiler chicken).

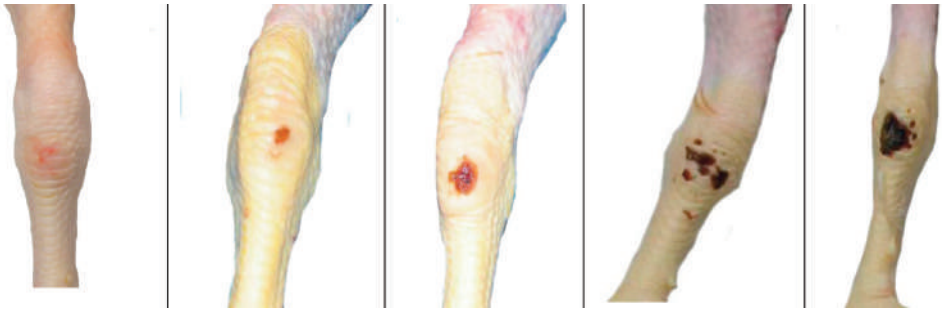
Welfare Criteria		Measures	
Good feeding	1	Absence of prolonged hunger	
	2	Absence of prolonged thirst	
Good housing	3	Comfort around resting	
	4	Thermal comfort	
	5	Ease of movement	
Good health	6	Absence of injuries	At farm – Lameness, hock burn, foot pad dermatitis At the slaughterhouse – Breast burns, hock burn, foot pad dermatitis
	7	Absence of disease	At farm – On farm mortality, culls on farm At the slaughterhouse – Ascites, dehydration, septicaemia, hepatitis, pericarditis, abscess
	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	Absence of fearfulness	

TABLE 4.5 Detailes of measures adopted in the 2009 WQ protocols for clinical scoring (poultry health status broiler chicken).

Lameness (gait score)	
Scope	Animal-based measure: Broiler chicken
Method description	Lameness is the inability to use one or more limbs in a normal manner. It can vary in severity from reduced ability or inability to bear weight, to total immobility. For all farm visits, which are made close to slaughter age, 150 birds approximately will be caught using a catching pen at random locations generated by computer. For very flighty bird (for example some free range bird) it may be necessary to catch small pens of birds. Each bird is individually encouraged to walk out of the pen and is scored as it does so. For each bird caught, the gait score will be recorded. The flock average gait score can be calculated by multiplying the number of birds in each gait score category, then dividing the total by the total number of birds scored – as shown in the table below. Birds are classified according to these criteria: 0. Normal, dextrous and agile. 1. Slight abnormality, but difficult to define. 2. Definite and identifiable abnormality. 3. Obvious abnormality, affects ability to move. 4. Severe abnormality, only takes a few steps. 5. Incapable of walking.
Classification	Individual level: Number of animals in each scoring class (0, 1, 2, 3, 4, 5) Percentage (%) of animals in each scoring class (0, 1, 2, 3, 4, 5).

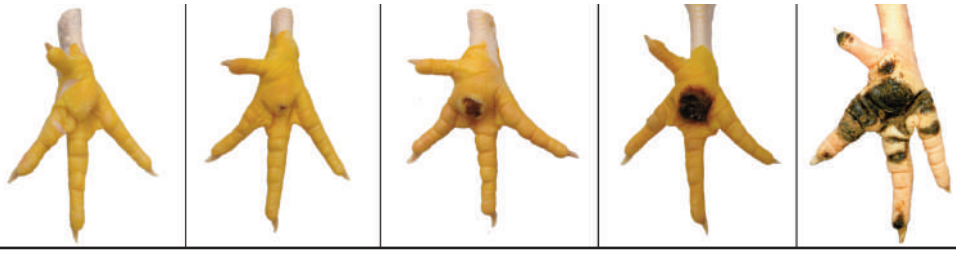
TABLE 4.5 CONT. Detailes of measures adopted in the 2009 WQ protocols for clinical scoring (poultry health status broiler chicken).

Hock burn	
Scope	Animal-based measure: Broiler chicken
Method description	Hock burn is a contact dermatitis found on the skin of the caudal (back) part of the hock joint. The skin is turned dark by contact with litter and consequently skin lesions can result. The scoring scale allows assessment of the severity of these lesions (see photographic reference). Assess the presence of hock burns with regard to the severity scale.
Classification	Individual level: Scoring categories a, b, c, d, e as photograph below. Number of animals in each scoring category a, b, c, d, e Combining categories 0 – No evidence of hock burn (score 'a') 1 – Evidence of superficial hock burn (score 'b' and 'c') – brown discolouration without scab. 2 – Evidence of severe hock burn (score 'd' and 'e') – black discoloura-tion with scab or ulcer.



1 2 3 4 5

Foot pad dermatitis	
Scope	Animal-based measure: Broiler chicken
Method description	Foot pad dermatitis (or pododermatitis) is a contact dermatitis found on the skin of the foot, most commonly on the central pad, but sometimes also on the toes. The skin is turned dark by contact with litter and consequently deep skin lesions can result. The scoring scale allows an assessment of the severity of these lesions (see photographic reference). Assess the presence of pododermatitis with regard to the severity scale. The foot pad needs to be clean before assessment. Scoring categories a,b,c,d,e as photograph below.
Classification	Individual level: Scoring categories a, b, c, d, e as photograph below. Number of animals in each scoring category a, b, c, d, e Combining categories 0 – No evidence of foot pad dermatitis (score 'a') 1 – Evidence of superficial foot pad dermatitis (score 'b' and 'c') - discolouration 2 – Evidence of severe foot pad dermatitis (score 'd' and 'e') – with scab or ulcer

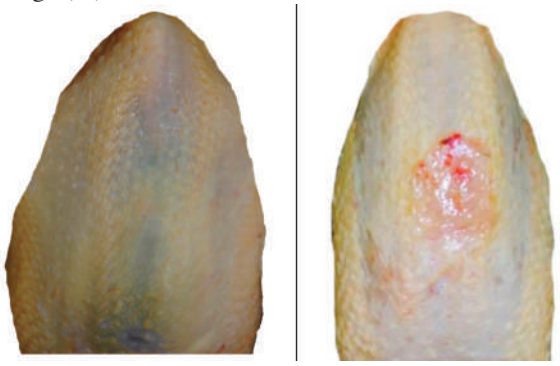


Source: (photos) © Colas, ITAVI (Institut technique de l'aviculture,France) (hock burn); © Butterworth, University of Bristol (foot pad dermatitis).

TABLE 4.5 CONT. Detailes of measures adopted in the 2009 WQ protocols for clinical scoring (poultry health status broiler chicken).

On-farm mortality	
Scope	Management-based measure: Broiler chicken
Method description	Mortality is defined as the 'uncontrolled' death of animals (as distinct from culling/euthanasia). The animals may die from, for example, heart failure, septicaemia, respiratory disease, acute infection or dehydration. Any animal which is 'found dead' on the floor in the house, or out on the field is considered a mortality. In the Additional animal unit information (see annex B), the animal unit manager is asked about mortality management on the farm based on data collected from farm records. Using house records of animal numbers placed, minus number died (but not including those actively culled, which are included in the measure 'culls on farm': Number of animals placed in house from the hatchery (A) Total number of animals which died and were found dead (but were not actively culled) in the last flock cycle (M). Calculate the percentage mortality using the following equation: Percentage of mortality = (M/A) x 100
Classification	Farm level: Percentage % of mortality on farm during the last flock cycle
Culls on farm	
Scope	Management-based measure: Broiler chicken
Method description	Culling is defined as birds which are actively and humanely killed by the animal unit manager for disease control purposes, or for lameness, sickness or disease – are know as 'culls'. In the Additional animal unit information (see annex B), the animal unit manager is asked about mortality management on the farm based on data collected from farm records. Using house records of bird numbers placed, minus those actively culled (but not including those found dead, which are included in the measure 'on-farm mortality': Number of birds placed in house from hatchery (A) Total number of birds which were actively culled (but not including those which died without being culled) during the flock cycle (C) Calculate the percentage culled using the following equation. Percentage of culling = (C/A) x 100 Culling is defined as birds which are actively and humanely killed by the animal unit manager for disease control purposes, or for lameness, sickness or disease – are know as 'culls'.
Classification	Percentage (%) culling.

TABLE 4.5 CONT. Detailes of measures adopted in the 2009 WQ protocols for clinical scoring (poultry health status broiler chicken).

Breast blister	
Scope	Animal-based measure: Broiler chicken (scope according to §5.1B.5)
Method description	Breast blisters are caused by a dermatitis of the skin overlying the keel (the central part of the breast area). The skin is softened and sometimes discoloured and may be infected and 'sticky,' or show as a raised blister. Observe birds on the line for 5 to 10 minutes. Doing so will provide a sample of n (line speed birds per minute (ls) x number of minutes (t)). Record number of birds passing per minute (line speed birds/min (ls)). Subsequently observe the birds where the breast is clearly visible after plucking. Count number of birds with breast blister lesions ('b') – a No evidence of breast blister b Evidence of breast blister To classify use calculation below, in which t = period of observation (minutes), B = number of birds with breast blister lesion, ls = line speed (birds per minute) and n = number of birds observed in total (t x ls). Percentage of birds with breast blister = (B / n) x 100%
Classification	Individual level: Percentage (%) of birds with breast blister
	
	a b
Ascites	
Scope	Animal-based measure: Broiler chicken (scope according to §5.1B.5)
Method description	Ascites is an accumulation of tissue fluid in the lungs, air sacs and abdomen resulting from cardiac insufficiency. Using data from slaughter house rejection records (collected for public health and disease monitoring purposes) it will be possible to determine how many birds were rejected for ascites from the total number slaugh-tered from the flock (n). Collect information from the meat inspection team at the slaughter house. They will reject birds with ascites as being unfit for human con-sumption and will record the number of birds rejected for ascites as they inspect them. Percentage of each category = (nx (where x = ascites)/total number slaughtered in group (n)) x 100(%)
Classification	Flock level: Percentage (%) of animals with ascites

Source: © Butterworth, University of Bristol (photo).

TABLE 4.5 CONT. Detailes of measures adopted in the 2009 WQ protocols for clinical scoring (poultry health status broiler chicken).

Dehydration	
Scope	Animal unit
Method description	Dehydration is a state in which the tissues of the bird become deficient in water – usually as a result of sickness making the bird unable to access water supplies, but also potentially (more rarely) due to failure of provision of water. Using data from slaughter house rejection records (collected for public health and disease monitoring purposes) it will be possible to determine how many birds were rejected for dehydration from the total number slaughtered from the flock (n). Collect information from the meat inspection team at the slaughter house. They will reject birds with dehydration as being unfit for human consumption and will record the number of birds rejected for dehydration as they inspect them. Percentage of each category = (nx (where x = dehydration)/total number slaughtered in group (n)) x 100%
Classification	Flock level: Percentage (%) of animals with dehydration
Septicaemia	
Scope	Animal-based measure: Broiler chicken (scope according to §5.1B.5)
Method description	Septicaemia is a generalized infection of the tissues of the bird following overwhelming infection. The internal organs become discoloured, with localized visible areas of infection. Using data from slaughter house rejection records (collected for public health and disease monitoring purposes) it will be possible to determine how many birds were rejected for septicaemia from the total number slaughtered from the flock (n). Collect information from the meat inspection team at the slaughter house. They will reject birds with septicaemia as being unfit for human consumption and will record the number of birds rejected for septicae-mia as they inspect them. Percentage of each category = (nx (where x = septicaemia)/total num-ber slaughtered in group (n)) x 100%
Classification	Flock level: Percentage (%) of animals with septicaemia
Hepatitis	
Scope	Animal-based measure: Broiler chicken (scope according to §5.1B.5)
Method description	Hepatitis is a localized infection of the liver, often with visible changes to the liver including discolouration, localized abcessation or infected areas and fibrin formation. Using data from slaughter house rejection records (collected for public health and disease monitoring purposes) it will be possible to determine how many birds were rejected for hepatitis from the total number slaughtered from the flock (n). Collect information from the meat inspection team at the slaughter house. They will reject birds with hepatitis as being unfit for human con-sumption and will record the number of birds rejected for hepatitis as they inspect them. Percentage of each category = (nx (where x = hepatitis)/total number slaughtered in group (n)) x 100%
Classification	Flock level: Percentage (%) of animals with hepatitis

TABLE 4.5 CONT. Detailes of measures adopted in the 2009 WQ protocols for clinical scoring (poultry health status broiler chicken).

Pericarditis	
Scope	Animal-based measure: Broiler chicken (scope according to §5.1B.5)
Method description	Pericarditis is an inflammation of the tissue which surrounds the heart – discoloration and thickening of this tissue (the pericardium) visible on the slaughter line, caused by infection or associated with ascites. Using data from slaughter house rejection records (collected for public health and disease monitoring purposes) it will be possible to determine how many birds were rejected for pericarditis from the total number slaughtered from the flock (n). Collect information from the meat inspection team at the slaughter house. They will reject birds with pericarditis as being unfit for human consumption and will record the number of birds rejected for pericarditis as they inspect them. Percentage of each category = (nx (where x – pericarditis)/total number slaughtered in group (n)) x 100%
Classification	Flock level: Percentage (%) of animals with pericarditis
Abscess (sub-cutaneous pus)	
Scope	Animal-based measure: Broiler chicken (scope according to §5.1B.5)
Method description	Abscess is a localized area of (usually bacterial) infection – resulting in tissue damage or necrosis and production of pus or inflammatory responses and visible as swollen areas, pus or areas of tissue discoloura-tion. Using data from slaughter house rejection records (collected for public health and disease monitoring purposes) it will be possible to determine how many birds were rejected for abscess from the total number slaughtered from the flock (n). Collect information from the meat inspection team at the slaughter house. They will reject birds with abscess as being unfit for human consumption and will record the number of birds rejected for abscess as they inspect them. Percentage of each category = (nx (where x - abscess)/total number slaughtered in group (n)) x 100%
Classification	Flock level: Percentage (%) of animals with abscess

TABLE 4.6 Measures adopted in the 2009 WQ protocols for clinical scoring (poultry health status laying hens).

Welfare Criteria		Measures	
Good feeding	1	Absence of prolonged hunger	
	2	Absence of prolonged thirst	
Good housing	3	Comfort around resting	At farm – Evidence of red mites
	4	Thermal comfort	
	5	Ease of movement	
Good health	6	Absence of injuries	At farm – Keel bone deformation, skin lesions, foot score, toe damage
	7	Absence of disease	At farm – On farm mortality, culls on farm, enlarged crops, eye pathologies, respiratory infections, enteritis, parasites, comb abnormalities
	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	Absence of fearfulness	

TABLE 4.7 Detailes of measures adopted in the 2009 WQ protocols for clinical scoring (poultry health status laying hens).

Evidence of Red Mites	
Scope	Animal- and resource-based measure: Laying hens
Method description	Examine both equipment in the house and birds for red mites (Dermanyssus Gallinae), on the most likely places, like: under perches, in cracks and crevices. See photographic reference. The animal unit manager may be of help to identify most likely places in the house. By scraping with a sharp device in cracks and crevices one can find red mites. Another way is to hold a white paper underneath a wire or perch and knock on it, so red mites fall onto the paper and can be seen. Severe infestations can be seen clearly as bunches of mites clutched to-gether. Also, severe infestations can be seen as blood spots on eggs. Furthermore inspect the birds for presence of red mites by inspecting comb, legs and breast skin. Also check dead birds if possible. Combine the findings of the inspection of the birds and the house into one score.
Classification	0 – No red mites detectable on birds and in the house 1 – Red mites found on birds or in the house, but not in large quantities and not clearly visible (e.g. no or minor numbers of mites found on hens and mites found in the house are hidden in cracks and crevices, not on many places and not in large quantities) 2 – Large quantities of red mites found on birds and/or in the house (e.g. mites can or can't be found on birds, but large amounts of mites are abundantly present on clearly visible spots)

TABLE 4.7 CONT. Detailes of measures adopted in the 2009 WQ protocols for clinical scoring (poultry health status laying hens).

Keel bone deformation	
Scope	Animal-based measure: Laying hens
Method description	Normal keel bones are straight without dips, bulbs, deviation or other abnormality. Abnormalities can be caused by badly shaped perches, healed breaks or decalcification of the keel. A keel bone deformation is any abnormality to the normal straight hape of the keel. Examine the breast of the hen by looking at it (in case of featherless breast) or by running fingers alongside and over the keel bone. Pick up a bird from within the penned group or from the litter or slatted floor, inspect visually and palpate the area. Compare to photographic reference.
Classification	Individual level: 0 – No deviations, deformations or thickened sections, keel bone completely strait 2 – Deviation of keel bone (incl. thickened sections) Percentage of birds in the flock in each category (0,2)

Straight keelbone | Keelbone with deviation | Keelbone with deformation

Skin lesions	
Scope	Animal-based measure: Laying hens
Method description	Skin lesions are wounds that have not yet been healed. Little wounds in a shape of punctiform pecks or scratches are not considered lesions, but if there are 3 or more pecks and/or scratches they are taken into account. Examine the rear and legs of the hen for presence of skin lesions. Pick up a bird from within the penned group or from the litter or slatted floor. Examine the rear and legs of the hen. Lift the feathers to examine the skin.
Classification	Individual level: 0 – No lesions, only single (<3) pecks (punctiform damage <0.5 cm diameter) or scratches 1 – At least one lesion <2 cm diameter at largest extent or ≥3 pecks or scratches 2 – At least one lesion ≥2 cm diameter at largest extent Percentage of birds in the flock in categories 0,1,2

TABLE 4.7 CONT. Detailes of measures adopted in the 2009 WQ protocols for clinical scoring (poultry health status laying hens).

Foot score	
Scope	Animal-based measure: Laying hens
Method description	The feet of hens should have a smooth skin without any wounds or abnormalities. Wire floors can cause hard patches or other proliferations of the epithelium. Inflammation of little skin damages of the foot can cause swelling of the foot, called bumble foot. This starts with minor swelling, but can get as bad as balloon-shaped feet. Although these inflammations can heal during the flock cycle, they are painful to the bird and therefore should be avoided as much as possible. The cause of bumble foot is not completely clear, but perch, hygiene and genotype may have influence. Examine both feet of the hen. Pick up a bird from within the penned group or from the litter or slatted floor. Examine both feet of the hen and choose the foot with the worst condition to score.
Classification	Individual level: 0 – Feet intact, no or minimal proliferation of epithelium 1 – Necrosis or proliferation of epithelium or chronic bumble foot with no to moderate swelling 2 – Swollen (dorsally visible). Percentage of the flock in each scoring category 0,1,2
Toe damage	
Scope	Animal-based measure: Laying hens
Method description	Due to bad design of equipment toes can be trapped and injured or torn off as a consequence. Toe damage is defined as wounds on one or more toes and/or missing (parts of) one or more toes. This measure considers a general impression of the flock relating to toe damage. 100 birds of the flock are examined individually, information from these birds will be included in the final score. During the flock work birds are visually inspected and an impression is obtained regarding the number of birds with toe damage. Classification reflects the number of birds with toe damage.
Classification	Flock level: 0 – No evidence of damaged toes 1 – Less than 3 birds with damaged toes 2 – Starting from 3 birds with damaged toes and more.

TABLE 4.7 CONT. Detailes of measures adopted in the 2009 WQ protocols for clinical scoring (poultry health status laying hens).

On-farm mortality	
Scope	Management-based measure: Laying hens
Method description	Mortality is defined as the 'uncontrolled' death of animals (as distinct from culling/euthanasia). The animals may die from, for example, septicaemia, respiratory disease, acute infection or dehydration. Any animal which is 'found dead' on the floor in the house, or out on the field is considered a mortality. In the Additional animal unit information (see annex B), the animal unit manager is asked about mortality management on the farm based on data collected from farm records. Using house records of animal numbers placed, died and culled: Number of animals placed in house from previous animal unit (A) Total number of animals which died and were found dead (but were not actively culled) during the last 12 months (M) If no information is specifically available on culled birds, simply use the flock records on mortality (which will then reflect mortality and culled together) Calculate the percentage mortality using the following equation: Percentage of mortality = (M/A) x 100
Classification	Percentage (%) of mortality on farm during the flock cycle.
Culls on farm	
Scope	Management-based measure: Laying hens
Method description	Culling is defined as birds which are actively and humanely killed by the animal unit manager for disease control purposes, or for lameness, sickness or disease – are know as 'culls'. Use farm records of the flock in the house to determine the number and percentage of culled birds. In the Additional animal unit information (see annex B), the animal unit manager is asked about mortality management on the farm based on data collected from farm records. Estimate the number of birds that are culled by checking the farm records. If no written information or accurate information of the animal unit manager is available, mark this issue as not known. In that situation culls will be included in the mortality figure. The percentage of birds culled can be calculated by dividing the number of culled birds by the number of birds housed. If culling is not specifically recorded then this should be noted – as this implies that all deaths are 'uncontrolled'.
Classification	Percentage (%) culling per flock cycle.
Enlarged crops	
Scope	Animal-based measure: Laying hens
Method description	This measure considers a general impression of the flock relating to enlarged crops. An enlarged crop is a condition in which the crop becomes distended with fluid and decomposing food. This abnormal de-velopment of the crop is usually visible as a pronounced swelling on the lower neck of the bird. 100 birds of the flock are examined individually; information from these birds will be included in the final score. During the flock work birds are visually inspected and an impression is obtained regarding the number of birds with enlarged crops. Classification reflects the number of birds with eye pathologies.
Classification	Flock level: 0 – No evidence of enlarged crops 1 – Less than 3 birds with enlarged crops 2 – Starting from 3 birds with enlarged crops and more

TABLE 4.7 CONT. Detailes of measures adopted in the 2009 WQ protocols for clinical scoring (poultry health status laying hens).

Eye pathologies	
Scope	Animal-based measure: Laying hens
Method description	This measure considers a general impression of the flock relating to eye pathologies. Eye pathologies is the swelling of the eyelids and skin around the eyes, closure of the eyes (or single eye) or discharge from the eyes. 100 birds of the flock are examined individually; information from these birds will be included in the final score. During the flock work birds are visually inspected and an impression is obtained regarding the number of birds with eye pathologies. Classification reflects the number of birds with eye pathologies.
Classification	Flock level: 0 – No evidence of eye pathologies 1 – Less that 3 birds with eye pathologies 2 – starting from 3 birds with eye pathologies and more
Respiratory infections	
Scope	Animal-based measure: Laying hens
Method description	This measure considers a general impression of the flock relating to respiratory infections. Respiratory infections are causing increased or laboured respiratory effort, sneezing and often associated with audible breathing sounds. 100 birds of the flock are examined individually, information from these birds will be included in the final score. During the flock work birds are visually inspected and an impression is obtained regarding the number of birds with respiratory infections. Classification reflects the number of birds with respiratory infections.
Classification	Flock level: 0 – No evidence of respiratory infections 1 – Less than 3 birds with respiratory infections 2 – Starting from 3 birds with respiratory infections and more
Enteritis	
Scope	Animal-based measure: Laying hens
Method description	This measure considers a general impression of the flock relating to enteritis. Enteritis are gut infections or digestive metabolism abnormalities often resulting in altered faecal state – discoloured faeces or in-creased liquid content or diarrhoea. 100 birds of the flock are examined individually; information from these birds will be included in the final score. During the flock work birds are visually inspected and an impression is obtained regarding the number of birds with enteritis. Classification reflects the number of birds with enteritis.
Classification	Flock level: 0 – No evidence of enteritis 1 – Less than 3 birds with enteritis 2 – Starting from 3 birds with enteritis and more

TABLE 4.7 CONT. Detailes of measures adopted in the 2009 WQ protocols for clinical scoring (poultry health status laying hens).

Parasites	
Scope	Animal- and management-based measure: Laying hens
Method description	Poultry species are susceptible to several parasites, incl. lice, mites, ticks, fleas, and intestinal worms. Parasites can live on birds (like various mites and lice) and can then be seen if the feather cover is inspected and moved aside by hand. Parasites can also live inside the hen (intestinal worms) and then mostly are not visible. One can suspect the presence of worms if hens are pale and week. Inspection of manure can give proof of presence of worms. Parasites that live in the envi-ronment of hens may be harmful as they transmit disease. This measure considers a general impression of the flock relating to parasites. Red mites are not considered here, as they are considered separately. Examine the henhouse and housing system. Evidence of fleas can be visible on windows or doors, where they leave their mesh. Beetles can be found in manure pits. Check manure belts to see if manure contains large amounts of worms. 100 birds of the flock are examined individually; information from these birds will be included in the final score. By inspection of naked body parts (comb) and by pushing feather cover aside and inspect the breast and legs, lice and mites will be detected if present. Examine the house and housing system for any evidence of parasites.
Classification	Flock level: 0 – Fly mesh on windows & doors 2 – No fly mesh on windows & doors and 0 – No evidence of parasites 2 – Evidence of parasites
Comb abnormalities	
Scope	Animal-based measure: Laying hens
Method description	A normal comb has an evenly red colour and no wounds or scratches. This measure considers a general impression of the flock relating to comb abnormalities. 100 birds of the flock are examined individually; information from these birds will be included in the final score. Apart from pecking wounds (that are scored separate) other comb wounds should be scored as well. Examples of comb abnormalities considered here can be found: • No blue or black areas should be present • Hens in high production may have a slightly paler comb, but too pale combs may indicate anaemia • If hens are dehydrated, combs may look more dried out and blue Comb abnormalities like pale combs or coloured spots are considered here, in which the classification reflects the number of birds with comb abnormalities.
Classification	Flock level: 0 – No evidence of comb abnormalities 1 – Less than 3 birds with comb abnormalities 2 – Starting from 3 birds with comb abnormalities and more.

Blue spot Blue area indicating dehydration

Source: © Fiks-van Niekerk, WUR (photo).

5

ASSESSMENT OF 'DEAD ON ARRIVAL' IN POULTRY

B. Algers, R. Westin, M. Raj and A. Velarde

5.1 SUMMARY

The measure of 'dead on arrival' (DOA) reflects catching, transport, lairage and animal health conditions. Current methods for assessment of DOA in poultry have been examined by questionnaires and visits to slaughter plants. The results show that most abattoirs already count all dead birds at shackling by counting every dead bird. The companies' own DOA records can thus be used for auditing. The most important factors believed to influence DOA are handling procedures during catching, compartment stocking density, transport time, lairage time and animal health within the assessed flocks.

5.2 INTRODUCTION

Birds that have died between catching and the moment of slaughter are termed 'dead on arrival' (DOA). Some identified factors associated with DOA percentage in broilers are ambient temperature, harvesting company, point of time for harvesting, breed, flock size, mean body weight, mean compartment stocking density, transport time, lairage time, and the interaction term transport x temperature (Nijdam et al., 2004). Estimates of the percentage of DOA:s vary from 0.05 to 0.057% (Nijdam et al., 2006). The large numbers of slaughtered poultry cause the estimated absolute number of birds dying during transport and lairage each year in UK alone to exceed 1.8 million individuals (Warriss et al., 2005). This is a major welfare concern. If assessed properly the percentage of DOA is proposed to be a valid measure of transport and lairage mortality (Weeks and Butterworth, 2004). The aim of this study was to develop and evaluate a method for assessment of the number of birds dead on arrival.

5.3 METHODS

Questionnaires regarding current methods for assessment of 'dead on arrival' have been sent to 9 abattoirs (3 in Spain, 3 in Sweden, 3 in United Kingdom). In addition, visits have been made to 3 of the responding abattoirs where unloading and routines for assessment of DOA was observed.

5.4 RESULTS

Answers were received from 6 of the 9 abattoirs. 3 did not want to participate. In all 6 responding abattoirs records of DOA:s are documented for each slaughtered flock. In 5 abattoirs, the assessment of DOA is done by exact counting of all dead birds in a flock. In one abattoir, assessment by counting the exact number of dead birds is only done if a truck driver suspects a high mortality rate. In all abattoirs dead birds are collected in a designated container during shackling. When shackling of one flock is finished, the container is emptied and the exact number of dead birds is counted (or estimated) and the sum is documented. In one abattoir, the personnel facilitate counting by using a mechanical counting device placed in the shackling area. When one dead bird is found and placed in the DOA container, the personnel push a button making the counting device increase one unit. When one flock has passed shackling, the obtained number of DOA:s is recorded and the counting device is reset. The abattoir that usually didn't count dead bird, assessed DOA by estimating the volume of dead birds in the DOA container. According to the slaughter house praxis, the personnel working with shackling are however supposed to always count all dead birds but this is not followed. At all visited abattoirs, assessment was done by counting. The staff was familiar with the assessment routines and did not think of this as difficult or time consuming. The mechanical counting device was thought to be very handy.

5.5 DISCUSSION

All asked companies kept their own records of DOA and assessment was done in very similar ways. Assessment was performed at shackling in all abattoirs and in 5 out of 6 the number of dead birds in each flock was counted. Only one company estimated the level of DOA based on volume in the DOA-container but this was not in line with the company's

praxis. Since most companies seem to assess DOA in slaughtered flocks by counting dead birds at shackling we conclude that this should be the standard method for assessment of DOA in poultry. In slaughter plants where this praxis is not followed, a mechanical counting device can facilitate the counting.

Several factors associated with DOA are related to the conditions during catching, transport and lairage. The results from a large study (Nijdam et al., 2004), including 1 907 broiler flocks, shows that the most important factors influencing DOA percentage are compartment stocking density (OR=1.09 for each additional bird in compartment), transport time (OR=1.06 for each additional 15 min), and lairage time (OR=1.3 for each additional 15 minutes). Handling during catching is also believed to be an important factor (Nijdam et al., 2006; Gregory, 1998). The level of DOA will therefore in particular reflect the condition of these parameters. However, pathological studies of DOA birds state that a high amount of birds are weekend due to infectious diseases as well as heart and circulation disorders (Nijdam et al., 2006; Gregory, 1998). The percentage DOA will therefore not only be an indicator of catching, transport and lairage conditions but also reflect the animal health within the assessed flock.

5.6 CONCLUSIONS

Assessment of 'dead on arrival' in poultry should be done at shackling by counting of every dead bird. A mechanical counting device can facilitate counting. The measure of DOA will in particular reflect handling procedures during catching, compartment stocking density, transport time, lairage time and animal health within the assessed flocks.

6

ASSESSMENT OF PRE-STUNNING SHOCK AND STUNNING EFFECTIVENESS

B. Algers, R. Westin and M. Raj

6.1 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 pre-stunning shocks and stunning effectiveness in poultry has been tested and evaluated and the general conclusion is that this is a very promising method.

Birds receiving pre-stunning shocks and birds showing signs of insufficient stunning are individually counted during one-minute intervals. Monitoring of pre-stunning shocks (electrical stunning only) is performed at the stunner entrance. Birds that flap their wings or climb the shackle are regarded as receivers of pre-stun shocks and are recorded. Counting is repeated at lest 10 times at each sample location (minimum of 300 individuals). Monitoring of stunning effectiveness is performed at 4 different locations: 1) at stunner exit, 2) at sticking point, 3) during bleeding and 4) at entrance to scalding tank. Birds that vocalize or show any movement (for example wing flapping, gasping for air and rising of the head) are regarded as 'insufficiently stunned' and recorded. In addition to scoring of individual birds, the main posture of the birds at stunner exit is evaluated according to the stunning method used and the applied currents/gas mixtures. Counting is repeated at lest 10 times at each sample location (minimum of 300 individuals)

The final scores have to be related to the speed of the slaughter line (number of slaughtered birds per minute) in order to be able to calculate the average proportion of birds receiving pre-stun shocks or insufficiently stunned. If the speed is not known, the number of slaughtered birds during one minute has to be counted as well.

6.2 INTRODUCTION

Stunning and slaughter of poultry has long been an area of great concern from a welfare point of view (Weeks and Butterworth, 2004). Uncrating and shackling of live birds are likely to impose a considerable stress on the birds as well as receiving pre-stunning electrical shocks (Raj, 1998; Weeks and Butterworth, 2004). Pre-stun shocks occur when the animal makes contact with the electrified water bath before it is stunned (Weeks and Butterworth, 2004). This results in wing flapping and/or high pitched vocalizations at stunner entrance (Wenzlawowicz and Holleben, 2001). Another problem area of welfare concern is that stunning isn't always completely effective (Raj, 1998). According to EU legislation methods used for stunning shall immediately render the animal unconscious until death by bleeding supervenes (Directive 93/119/EC). Since many years, electrical water bath stunners is the most frequently used method for stunning of poultry under commercial conditions (Raj, 1998; Bilgili, 1999). However, during the recent years stunning methods using carbon dioxide or other gas mixtures has been developed and are now commercially available (Gerrizen et al., 2000; van der Sluis, 2005). The basic principle involved in electrical stunning is that an electric current is passed through the brain to induce epilepsy which is accompanied by unconsciousness (Raj, 1998). The electric stimulation will also cause an instantaneous depolarization (contraction) of skeletal muscle, making the bird unable to move (Wenzlawowicz and Holleben, 2001). The stunning is however reversible and if cardiac arrest isn't induced, the bird may regain consciousness if the cut or bleeding is incomplete (Raj, 1998; Bilgili,1999). In systems based on gas stunning, birds tend to regain consciousness rapidly on exit from the gaseous atmosphere unless they have been killed by the gas (Raj, 1998). Clinical assessment of stun quality should therefore include monitoring of animals after as well as before cutting. A checklist for assessment of the stunning process in poultry has been developed by Shütt and Abraham (in: Wenzlawowicz and Holleben, 2001). This is listed in Box 6.1.

Slaughter lines in Europe today commonly operate at speeds of 100 to 140 birds per minute (Bilgili, 1999). There are reasons to believe that some of these proposed signs will be very difficult to detect during the regular slaughter process due to high speed of the shackle line. The aim of this study is to develop and evaluate a method for clinical assessment of pre-stun shocks and stunning effectiveness in poultry possible to conduct under commercial conditions.

Box 6.1 Checklist for assessment of the stunning process.

A	At stunner entrance (Immediate stun?)
	Heads properly immersed?
	No wing flapping, no vocalization?
	No premature electrical shocks?
	Instantaneous onset of tonic spasm with current contact?
B	At stunner exit (Efficient stun?)
	No birds 'climbing their shackles'?
	No wing flapping, no vocalization?
	No respiration?
	No corneal reflex?
	Birds still in a tonic state (with current flow durations <8 sec) respectively fully relaxed (with current flow duration > 10 sec and cardiac arrest)?
C	At sticking point (Lasting stun?)
	No return of respiration?
	No return of corneal reflex?
	No head raising and wing flapping triggered by sticking?
	No reaction to painful stimuli?
	Bleeding rapid and profuse?
D	During bleeding (No signs of return of consciousness?)
	No return of regular breeding?
	No return of spontaneous blinking?
	No raising of heads?
	No reaction to painful stimuli?
	No hypoxemic convulsions at the end of bleed-out (cardiac arrest stunning only!)?
E	At entrance to scalding tank (Confirmation of death?)
	No respiration?
	No corneal reflex?
	No muscle tone/no movements?
	No reaction to painful stimuli?
	No reaction to contact with scalding water?
F	After scalding and plucking (Birds properly bled?)
	All necks properly cut?

Source: Wenzlawowicz and Holleben, 2001.

6.3 METHODS

Visits have been made to 4 abattoirs (2 high speed, electrical stunning, 1 low speed electrical stunning, 1 high and low speed gas stunning). During each visit the personnel were asked about their regular procedures for observation of stunning effectiveness. In 3 of the abattoirs birds were filmed at 5 different locations (at stunner entrance, at stunner exit, at sticking point, during bleeding and at entrance to scalding tank). Based on the checklist for assessment of poultry stunning process (see Box 6.1) proposed by Shütt and Abraham (in: Wenzlawowicz and Holleben, 2001) and impressions from the slaughter plant visits, a method for assessment of pre-stun shocks and stun quality was developed. (For details, see chapter 3). From collected film material a film (film A) exemplifying signs of pre-stun shocks and sufficient and insufficient stunning was produced. Another film (film B), consisting of 1 minute interval sequences (20+20 sequences filmed at stunner

entrance, 10 sequences each filmed at stunner exit, at sticking point, during bleeding and at entrance to scalding tank) was produced as material for testing of inter-observer reliability.

For evaluation of the method and testing of inter-observer reliability, film A was shown to 3 observers and a discussion about how to monitor pre stun shocks and stunning effectiveness was held with each observer. Following this, the observers individually monitored every sequence in film B. 20 one-minute sequences (10 from high speed line,100 birds/minute; 10 from low speed line, 30 birds/minute) where monitored twice for occurrence of pre stun shocks but in reversed order. First time, wing flapping and shackle climbing birds were counted during each interval. The second time vocalizing birds were counted. Monitoring of stunning effectiveness was performed by counting birds showing signs of insufficient stunning (vocalization, wing flapping, gasping for air, raising head or other movement) during 10 one-minute sequences from each of 4 different locations (at stunner exit, at sticking point, during bleeding and at entrance to scalding tank).

All figures obtained from low speed lines (30 birds per minute) were multiplied with 3.3 to be able to be equivalent to the figures obtained from high speed lines (100 birds per minute), representing number of birds showing signs of pre-stun shock or insufficient stunning per 100 birds. All statistical analysis was than performed in Minitab (MINI-TAB® Release 14.1, Minitab Inc, 2003). The combined monitoring results of all three observers for the 10 low speed sequences filmed at stunner entrance, was statistically analysed for validation of 'count of wing flapping/shackle climbing' versus 'count of vocalizations' as equal methods for monitoring of pre stun shocks. The scoring means of the two methods were compared by analysis of Pearson's correlation coefficient (Dalin, 2003).

Statistical analysis of inter-observer repeatability was performed by calculation of Spearman's correlation over all pairs of observers in each location as proposed by Engel and van Reenen (2006). Counting of vocalizations in the high speed abattoir (10 sequences) was impossible due to high noise level from ventilating fans and vocalizations from shackled birds so these sequences were not included in the statistical analysis. No monitored animal displayed any signs of efficient stun at stunner exit so no statistical analysis was performed for this location.

6.4 RESULTS

Mean scoring results for monitoring of all locations are listed in Table 6.1.

The average mean scores for pre stun shock based on wing flapping from the low speed line only was 27.5 and the average score based on counts of vocalizing birds was 30.8/100 birds. Comparison of occurrence of pre stunning shock based on counts of wing

flapping/shackle climbing versus counts of vocalisations by calculation of Pearson's correlation showed that there was no significant difference between the two methods (Pearson's correlation = 0.051, p-value =0,790).

Spearman's correlation coefficients over all pairs of observers are listed in Table 6.2. The average Spearman correlation coefficient was >0.80 for signs of insufficient stunning at sticking point, during bleeding and when counting wing flapping as a sign of pre stunning shocks indicating a strong inter-observer agreement at these locations.

6.5 DISCUSSION

The high speed of today's slaughter lines makes monitoring of several important behavioural signs of a returning consciousness such as return of breathing very difficult. To be possible to perform monitoring under commercial conditions, this method is only based on behaviours possible to detect when standing aside of the slaughter line (wing flapping, gasping for air, posture etc.). But, occurrence of these behaviours does not necessarily mean that the bird is conscious. To be able to decide if a bird is conscious or not more sophisticated observation is needed such as control of corneal and comb reflex (Wenzlawowicz and Holleben, 2001). As long as a bird is unconscious the stunning is sufficient and to say that this scoring method measures the 'true' value of insufficiently

TABLE 6.1 Mean scores of occurrence of pre stun shocks and signs of insufficient stunning per 100 birds monitored by three different observers.

Pair of Observers	Pre-stun shock		Stunning effectiveness		
	Wing flapping (n=20)	Vocalization (n=10)	At sticking point (n=10)	During bleeding (n=10)	At entrance to scalding tank (n=10)
A	15.5	28.0	.8	.9	.3
B	17.1	33.7	.7	1.1	.1
C	12.5	30.6	.9	.8	.3

TABLE 6.2 Spearman's correlation coefficients for evaluation of inter-observer reliability between 3 different observers.

Pair of Observers	Pre-stun shock		Stunning effectiveness		
	Wing flapping (n=20)	Vocalization (n=10)	At sticking point (n=10)	During bleeding (n=10)	At entrance to scalding tank (n=10)
A–B	.866	.635	.996	.848	.509
A–C	.963	.823	.996	1.000	1.000
B–C	.819	.532	.984	.848	.509
Average of all pairs	.883	.663	.992	.899	.673

stunned birds is therefore in a way incorrect. However, if we regard a bird showing signs of movement such as wing flapping as a 'high risk individual' for return of consciousness we believe that the proportion of 'high risk individuals' will be closely related to the proportion of conscious animals.

The results show that on average there is a high inter-observer agreement for the conducted method. Spearman's correlation provides a robust measure of the similarity between the scores of different observers. A 100% agreement between observers would give a correlation value of 1. Correlations >0.80 indicate a strong inter-rater agreement (Engel & van Reenen, 2006). The average Spearman correlation for stunning effectiveness at entrance to scalding tank is only 0.67 which points at a much weaker agreement between the observers at this location. But, looking at the correlation between each pair of observers we see that between observer A–C the agreement is perfect (1.00) but for observer A–B and B–C as low as 0.51. From the mean scores in Table 6.2 we can conclude that observer A and C detected 3 insufficiently stunned birds out of 1000 (0,003%) and observer B detected 1 out of 1000 birds (0,001%). Since so few birds are showing signs of insufficient stunning at this location, one missed individual has great effect on the Spearman correlation. In spite of a low average agreement according to Spearman's correlation at entrance to scalding tank, we evaluate inter-observer agreement as good at this location.

The result from 10 film sequences were excluded from the statistical analysis since high noise levels from fans and shackled birds made monitoring of vocalizations impossible. The average value of Spearman's correlation for pre stun shocks based on vocalizing birds was lower than for wing flapping birds (0.66 versus 0.88). This indicates that vocalization is more difficult to assess than wing flapping. Comparison of the total mean scores of pre stun shock based on vocalizations contra wing flapping/shackle climbing showed that there was no statistical difference. This means that both methods give the same result. Because of these facts, we conclude that counting of vocalizing birds at stunner can be excluded from the monitoring procedures.

Electrical stunning in poultry will induce a depolarization of skeletal muscle, but not necessarily unconsciousness (Wenzlawowicz and Holleben, 2001). Muscle depolarization causes muscle contraction which makes the bird rigid and unable to move. Muscle depolarisation without loss of consciousness means that fully conscious and sensible birds that appear to be properly stunned are being cut and bleed. From welfare point of view, minimum current required per bird in the water bath stunner can be determined using three criteria mention by Raj (1998): 1) induction of epileptiform activity in the brain, 2) abolition of somatosensory evoked potentials (SEP) in the brain, and 3) induction of cardiac arrest at stunning. None of these criteria are possible to clinically assess during commercial slaughter conditions (Wenzlawowicz and Holleben, 2001). It is therefore strongly recommended that evaluation of the electrotechnical conditions (voltage, amperage, frequency, waveform and exposure duration) is performed in addition to the clinical assessments. Monitoring the behaviour without knowledge of the applied currents can otherwise lead to a false interpretation of the stunning effectiveness (Wenzlawowicz and Holleben, 2001).

6.6 CONCLUSIONS

The presented method is a good candidate for measuring of stunning effectiveness in poultry under commercial conditions. However, counting of vocalizations as a sign of pre stun shocks should be excluded since the noise level at a slaughter plant makes auditing difficult or impossible. Evaluation of the electrotechnical conditions (voltage, amperage, frequency, waveform and exposure duration) must be performed in addition to clinical assessments.

7

ASSESSMENT OF THE HUMAN–ANIMAL RELATIONSHIP IN LAYING HENS

C. Graml, K. Niebuhr and S. Waiblinger

7.1 SUMMARY

While for caged hens validated tests assessing the human–animal relationship (HAR) are described in the literature (e.g. Barnett et al., 1992; Jones, 1996), this does not apply for non-cage systems with large group sizes. The aim of the study was the validation of several tests developed to assess HAR in non-caged laying hens. Additionally we tried to calibrate the tests for cage and non-cage systems against each other by direct comparison of measures of the same type or by use of a test in a standardized environment (arena test after Barnett et al., 1992), performed in the same way in both systems.

Based on a literature survey we developed or modified 3 tests to assess HAR in non-caged laying hens: avoidance distance test (ADT), touch test (TT) and stationary person test (SPT). Validity was tested by an on farm survey on 50 non-cage flocks (Niebuhr et al., 2007) and an experimental study in 12 groups of laying hens with two different handling treatments (Graml et al., 2008a).

The developed tests for non-cage systems turned out to be fast and easy to perform on farm and a good validity was shown. Measures of ADT and TT showed significant and moderate to mostly high correlations between each other (r_s = 0.5 to 0.8) and lower correlations ($r_s \leq 0.4$) with a test for general fearfulness (novel object test), indicating good specificity for measuring the hen's relationship to humans. In an on farm survey associations to stockpersons' attitude and behaviour further supported the validity. In an experimental study exposure to additional positive human contact lead to a significant decrease of the avoidance distance (p=0.004) whereas the number of touched birds in TT increased (p=0.002).

Within cage systems the tests described in the literature had to be modified for application on commercial farm. These modified tests were easy to perform on farm and showed

significant correlations (r_s=0.5–0.7) between the different tests used to assess HAR, suggesting a good validity.

The calibration between the tests for cage and non-cage systems was studied in 14 units with cage systems (10 with conventional cages and 4 with enriched cages) and 10 flocks housed in free range systems (Graml et al., 2008b). The calibration failed due to a lack of directly comparable results and a lack of correlations to the standardized arena test within cage systems. For non-cage systems we found significant correlations (r_s>0.6) between the tests assessing HAR in the flock and the arena test. At the moment, the comparison of results seems only possible by using the ADT in both systems and creating a parameter which takes the system into account.

For practical application in an on-farm scheme to assess animal welfare we recommend the avoidance distance test as most promising measure for both systems due to the high validity and possibility to compare between cage and non-cage systems. Nevertheless we see the touch test as a promising tool to assess HAR in non-cage systems that merit further evaluation. In sight of the large variety of aviary and single tier systems currently in use throughout Europe testing hens in the litter area could be more suitable. Additionally in this test an increase of sample size requires comparatively less time.

7.2 INTRODUCTION

While for caged hens validated tests for assessing the human-animal relationship (HAR) are described (e.g. Barnett et al., 1992; Jones, 1996), this does not apply for non-cage systems with large group sizes. Based on a literature survey we developed or modified 3 tests to assess the human-animal-relationship (HAR) in non-caged laying hens. In a pre-study these tests showed a high test-retest repeatability (r_s>0.8) when used in 6 free-range flocks (Raubek et al., 2007).

Measures used to assess HAR should be tested for validity prior to their use which was the aim of this study. We studied these developed tests, supposed to measure HAR in laying hens, by performing an on farm survey (Niebuhr et al., 2007) and an experimental study (Graml et al., 2008a). Following Waiblinger et al. (2006), conceptually related measures should empirically be associated with one another (i.e. convergent validity), whereas conceptually unrelated measures should be empirically independent (i.e. discriminant validity). According to this concept we would expect high correlations between the tests for assessment of HAR and lower correlations with a test for general fearfulness indicating high specificity of the test. We investigated this in a study on 50 flocks housed in non-cage systems. Additionally, in an experimental study we exposed 12 groups of laying hens to two different treatments. Assuming that additional positive human contact reduces the fear

of humans and improves the hens' relationship to humans, we expected the influence of the different treatments to be reflected in the results of the tests.

Cage and non-cage systems differ largely, and therefore it is difficult to use the same method and tests to assess HAR in both housing systems. However, the aim of Welfare Quality is to develop a welfare assessment scheme for all housing systems of laying hens and to make results comparable. To obtain comparable results between the two housing systems two possibilities seem to exist. We could either directly compare the results of the tests, if they are similar enough, and produce a measure of the same type (e.g. avoidance distance). Or calibrate the different measures against each other by use of a standardized test, performed in the same way in both systems. We investigated both approaches in 14 flocks housed in cages and 10 free-range flocks on commercial laying hen farms (Graml et al., 2008b).

7.3 METHODS

7.3.1 Validity of the Tests

An on farm survey was conducted on 50 flocks with brown hybrids housed in non-cage systems (deep litter, free-range). Flock size varied between 500–3000 birds per flock and the birds were between 30 to 35 weeks of age. Tests included the reactions of hens towards a stationary person (SPT), a moving human approaching individual hens (ADT) and a stationary human trying to touch individual hens (TT; Table 7.1).

Besides the 3 tests for assessing the human-animal-relationship (HAR) we performed one test to measure general fearfulness (novel object test; NOT). As novel object we used a colourful ball that was placed in the litter area. The test person filmed a defined area (1 x 0.8 m) around the ball from a distance of 1.5 m. The video was later analyzed by scan sampling in 10 s intervals over a 2-minute period. The test was performed twice per flock and the total number of hens in the defined area was calculated for the flock. The relationships between farm factors, stockperson behaviour and attitudes and the results of the tests were analysed by multivariate analysis.

In an experimental study we exposed 6 groups of laying hens to additional, positive human contact while 6 control groups received only minimal, neutral human contact. Birds were 25 and 31 weeks of age when starting the treatment and each group comprised about 500 brown hybrids. Immediately before and after a 2-week treatment period, all 12 groups were tested with the 3 tests to be validated. For details on treatments and study design, see Graml et al. (2008).

TABLE 7.1 Tests to assess HAR in non-cage systems.

Test	Description
SPT	= stationary person test Test person is standing with the back to the wall in the litter area and films a defined area (1 x 0.8 m) over a period of 2 minutes. The video camera is held at the height of the test person's head. The test is performed twice per flock and the video tape later analyzed by scan sampling. In 10 s intervals the number of hens (all visible parts) per area are counted and averaged over the flock. measure: • (SPT) average number of hens per defined area
TT	= touch test The test person approaches a group of at least 3 hens in the litter area, squats for 10 seconds and then counts the number of hens in arm length. Afterwards any 3 hens of those in reach are tried to be touched one by one. Every attempt to approach a group of birds is considered as a trial even if all hens from the group withdraw from the approaching or squatting test person. The procedure is repeated until 21 hens could be touched or until 21 trials were performed, independent of the number of touched birds (minimum 0 hens touched; maximum 23 hens touched in 21 trials). measures: • (TT reach) total number of hens in reach per trial • (TT touch) total number of hens touched per trial (max. 3)
ADT	= avoidance distance test The test person walks parallel and at a distance of 1.5 m to the slatted area. After selecting any hen on the edge of the dropping pit, the test person turns towards the hen, approaches the hen in a speed of 1 step/sec until it withdraws and then measures the distance from the test person's hand to the position of the hen's feet before withdrawal (5 cm resolution; folding rule). The hen's reaction is considered as withdrawal when the bird is lifting its second foot, meaning that the bird really moves away. Changes in the body axis, turning away of the head, lifting and staying on one foot are ignored. If the bird does not withdraw and can be touched by the approaching human, the avoidance distances is 0 cm. A total of 21 hens per flock are tested and the median per flock is calculated. measures: • (ADTmed) median avoidance distance measured in 5 cm resolutions per flock (avoidance distance = 0-150 cm) • (ADTtouch) percentage of birds with avoidance distance = 0 cm (bird is touched)

7.3.2 CALIBRATION BETWEEN SYSTEMS

Tests in Cages

For the calibration of the tests between cage and non-cage systems we tested 14 flocks in cages (10 in conventional and 4 in furnished cages) and 10 flocks in non-cage systems with the tests for each system and one standardized arena test (after Barnett et al., 1992), for both systems. All tests were conducted on brown hybrids at ages between 20 and 70 weeks. In non-cage systems flock size varied between 500 and 3,000 birds; in cage systems the group size per cage varied between 3 to 60 birds per cage, flock size in cages varied between 1,500 and 24,000. For more information see Graml et al. (2008b).

For assessing HAR in cage systems we developed or modified 3 tests for application on hens staying inside the home cages (Table 7.2). Tests described in the literature had to be modified due to the following reasons:

- Varying group size from 3 to 6 hens in conventional, up to 60 birds in enriched cages
- The tests described in literature were only used in cages with single housed hens or with stable group sizes. Larger group sizes make it impossible to observe and follow the reaction (position, orientation and posture) of a single bird; as described in the literature (Barnett, 1989; Jones, 1985, 1993).
- Corridor width. The tests described in literature need large observation or approaching distances and cannot be performed in commercial farms where corridor width is usually 100 cm or less.

TABLE 7.2 Tests to assess HAR in cage systems.

Test	Description
HT	= husbandry test The test person walks down the corridor (hands by side, looking at the food trough to avoid direct eye contact) at 1 step/s and at a distance of 60 cm from the body to the front of the cage row. A cage in a minimum distance of 1.5 m and with at least 3 heads through the front wire-mesh is selected. While walking down the corridor, the test person counts the number of heads out of the selected cage at a distance of 1.5 m before reaching the edge of the cage (measure 1) and directly when passing the edge (measure 2). During walking further, the next cage is selected until 10 cages are tested. measure: • (HT) Average percentage of heads staying outside the cage when the test person is passing = relation of measure 2 (number of heads out when reaching the cage) to measure 1 (number of heads out in 1.5 m distance) * 100, (min 0; max 100)
SPT	= stationary person test The test person walks down the corridor and stops at a cage with at least 3 heads stretched through the front wire. Body is then orientated towards the front of the cage at a distance of 60 cm. Over the duration of 2 minutes all hens with their heads out of the cage and hens that could potentially stretch their head through the front wires are recorded at 10 s intervals. Hens are recorded as potentially head out when their head is estimated within the first 15 cm of the front part of the cage and their body axis is oriented at a maximum of 180 ° (parallel) to the front wire-mesh. 10 cages per flock are tested. measure: • (SPT) proportion of birds with head out of the cage to sum of birds with potentially head out (in %)
ADT	= avoidance distance test The test person approaches individual hens stretching their head out of the cage (head stretched through the front wire-mesh including it's comb) with one hand held 15 cm in front of the body starting the approach from a distance of 60 cm (from hand to front of the cage). The distance of 60 cm to the cage is measured before starting the test by use of a folding rule and then the test person walks parallel to the cages in this estimated distance. After selecting any hen with head out of the cage the test person turns frontal towards the hen and approaches the hen in a speed of 1 step/s until it withdraws it's head into the cage and then measures the distance from the test person's hand to the front wire-mesh (5 cm resolution; folding rule). measure: • (ADT) median avoidance distance in 5 cm resolutions per flock (max 60 cm)

We modified and developed 3 tests, which are independent of the number of hens in each cage. They mainly record the reactions of animals that stretch the head through the front wire-mesh when a test person is passing by, standing still in front of the cage or approaching individual hens. Thus, we avoid the problem with changing group sizes. To be considered as "head out", then hens' heads need to be stretched completely through the front wire-mesh including the comb. No bird was used in more than one test and in each row the test cages were a minimum of 1.5 m apart between consecutive tests. All tests were performed in the same tier level at breast height of the test person which corresponded with tier level 2 or 3 depending on the cage design. Locations of tests were evenly distributed throughout the house. The order of tests was pseudo-randomised and the different tests were performed in sequence along a row.

To obtain information about the discriminant validity of the tests we performed a test for the assessment of general fearfulness (novel object test) besides the 3 tests used to assess HAR. Again group size made the observation of single birds as described by Jones (1985) impossible. Also the proposed use of a brightly coloured wooden rod as a novel object turned out to be impracticable due to the intense avoidance reaction of the birds suggesting low variability between different flocks. Therefore we modified the test by using a stripe of white plastic (5 x 14 cm) vertically fixed in the food trough with a short wire. After placing the novel object in the food trough the test person withdrew to a distance of 1.5 m to the side of the cage. As described for the SPT in cage systems, all hens with head out of the cage and hens that could potentially stretch the head through the front wires were recorded over the duration of 2 minutes at 10 s intervals. The test was repeated in 10 cages and the proportion of hens with heads out of the cage per flock was calculated. Additionally the latency until the first hen stretched its head (including the comb) through the mesh after the fixation of the novel object was measured by use of a stop watch.

7.3.3 Tests in Non-cage Systems

To assess HAR in non-cage systems the SPT, ADT and TT were used (Table 7.1). Additionally a novel object test described before (see validity study) was used.

Arena Test for Both Systems

In order to compare and calibrate results between cage and non-cage systems we performed a standardized arena test (ST) besides the developed tests for the two housing systems. We decided to use the so called shute test described by Barnett et al. (1992). This test has already been used to assess fear of humans and can be applied in any housing system. To create equal conditions, the test arena should be an unusual, unknown surrounding for all test birds. Therefore, tests using cages (like the box plus test person test described by Jones et al., 1993) seemed not applicable.

The arena consisted of a table (0.8 m x 1.2 m, height from the floor 0.75 m) with 3 areas (each measuring 0.8 x 0.4 m) situated at the end of a corridor (length 2.4 m, width 0.8 m, height 1.8 m). A rough, structured paper covered the table to avoid a slippery surface. The corridor was a construct of beige curtains suspended from a plastic tube frame and one curtain was bending over the table at a height of 1.8 m. Thus, the birds were surrounded by curtains except in the direction towards the test person.

After performing the tests to assess HAR described above, we caught 30 hens, chosen randomly from the flock or cages. Birds were tested singly in the arena in a room adjacent to the hen house out of visible contact to their flockmates. Each hen was placed on the most forward of the three areas on the table. Birds were oriented towards the test person. The test person withdrew to a distance of 2.4 m from the edge of the table and slowly approached the animal after a break of 15 s. The approach was interrupted by breaks of 5 s, at 1.6, 0.8 and 0 m from the table. The bird's position and orientation were recorded after the breaks. Results were immediately written on a paper (A6) hidden in the left hand. To prevent repeated measures, birds were marked on their legs after being tested.

The following measures were analyzed:

1. proportion of birds that stayed in the most forward oriented area of the table (STstay);
2. proportion of birds that flew up or down from the table (STflyup);
3. proportion of birds orientated forward (STorientation).

7.4 RESULTS

7.4.1 Validity of the Tests

On-farm Survey

The tests in the non-cage housing systems turned out to be fast and easy to perform. As shown in Table 7.3, the tests showed significant and high correlations with the other tests for assessing human-animal relationship (HAR) and lower correlations with the test for general fearfulness (NOT). The ADTtouch showed no correlation and the remaining measures moderate correlations with the NOT. The ADTtouch showed also a lower correlation with the SPT than the other tests.

TABLE 7.3 Spearman correlation coefficients for several measures of the tests assessing HAR and one test for general fearfulness (NOT) conducted in 50 flocks of non-cage systems.

					Tests assessing HAR					Gen. fearfulness
Tests	measures	median	min–max	25%–75%	SPT	TT touch	TT reach	ADT med	ADT touch	NOT
SPT	Number of hens in defined litter area	3.31	.00–18.67	.42–6.17	1	.69 p=.000	.84 p=.000	–.44 p=.001	.31 p=.027	.43 p=.002
TT touch	Average number of touched hens per trial	.24	.00–3.00	.00–.81	.69 p=.000	1	.88 p=.000	–.71 p=.000	.60 p=.000	.42 p=.003
TT reach	Average number of hens in reach per trial	.67	.00–3.00	.036–1.87	.84 p=.000	.88 p=.000	1	–.60 p=.000	.52 p=.0000	.42 p=.003
ADT med	Median of distance (cm) when bird withdrew an appr. human	50.0	.00–150	30.0–72.50	–.44 p=.001	–.71 p=.000	–.60 p=.000	1	–.82 p=.000	–.39 p=.005
ADT touch	Prop. hens that could be touched by an appr. human	2.38	.00–66.67	.00–23.81	.31 p=.027	.60 p=.000	.52 p=.000	–.82 p=.000	1	.15 p=.313
NOT	Total number of hens on defined area around novel object	.74	.00–12.67	.05–5.34	.43 p=.002	.43 p=.002	.42 p=.003	–.39 p=.005	.15 p=.313	1

Experimental Study

Exposure to additional positive human contact led to significant differences between handling and control groups (Table 7.4). After a 2-week treatment period the avoidance distance decreased significantly within the handling groups whereas the proportion of touched birds increased (Figure 7.1).

7.4.2 CALIBRATION BETWEEN SYSTEMS

The arena test should be independent of the housing system. However, significantly more birds stayed orientated forward (p=0.003) to an approaching human in the cage systems. No differences could be found in the proportion of birds flying up or staying in the most forward area of the table between the two systems.

TABLE 7.4 Results of three tests to assess HAR before and after a two-week treatment period in two different handling groups (additional positive contact versus minimal neutral contact). Data were analyzed with Mann-Whitney-test for differences between groups and with Wilcoxon-test for differences within the same group.

			SPT	TT touch	TT reach	ADT med
Additional	before treatment	Median	12.06	2.12	3.12	35.00
		Min–max	9.67–16.25	1.24–2.88	1.65–6.00	25–40
		25%–75%	9.79–15.06	1.67–2.69	2.53–4.88	25.00–36.25
	after treatment	Median	2.83	3.0	7.93	12.50
		Min–max	17.5–23.5	.00–3.00	5.29–1.86	0–25
		25%–75%	17.91–22.62	3.00–3.00	6.25–9.79	.00–21.25
Minimal contact group	before treatment	Median	12.81	2.10	3.42	40
		Min–max	4.71–16.13	1.31–2.88	1.69–5.50	20–60
		25%–75%	8.77–15.78	1.54–2.55	1.86–4.54	31.25–45.00
	after treatment	Median	11.77	1.76	2.43	37.50
		Min–max	5.8–17.2	.95–2.10	1.33–3.00	35–50
		25%–75%	9.05–15.46	1.22–1.96	1.70–2.83	35.00–46.25
Differences between groups before treatment			p=.749	p=.810	p=.748	p=.183
Differences between groups after treatment			p=.004	p=.002	p=.004	p=.004
Changes within additional contact group			p=.028	p=.028	p=.028	p=.028
Changes within minimal contact group			p=.600	p=.046	p=.075	p=.786

Note: Data were analyzed with Mann-Whitney-test for differences between groups and with Wilcoxon-test for differences within the same group.

An age dependency of the tests assessing HAR in non-cage systems was found as shown in Table 7.7. In cage systems only the SPT and TT tend to correlate with the age of the hens.

With respect to inter-test correlations, for 10 flocks housed in non-cage systems, the tests showed high correlations between all measures (Table 7.5). The standardized arena test (ST) showed significant correlations with the measures assessing HAR but not with NOT (general fearfulness). No correlations at all could be found for STorientation.

For the 14 flocks housed in cage systems tests also showed good correlations between all measures assessing HAR and lower correlations with the test for general fearfulness being significant only for SPT (Table 7.6). However, correlations are lower than in non-cage systems. The HT showed lower correlations than the other tests. In contrast to the non-cage systems, no significant correlations could be found between the tests assessing HAR and the standardized arena test. Only results of the husbandry test tend to correlate with the proportion of birds that stayed in the most forward oriented area of the table (STstay) as does the latency of the first head coming out of the cage after placing a novel object into the food trough (NOTlat) with the birds that flew in the arena test.

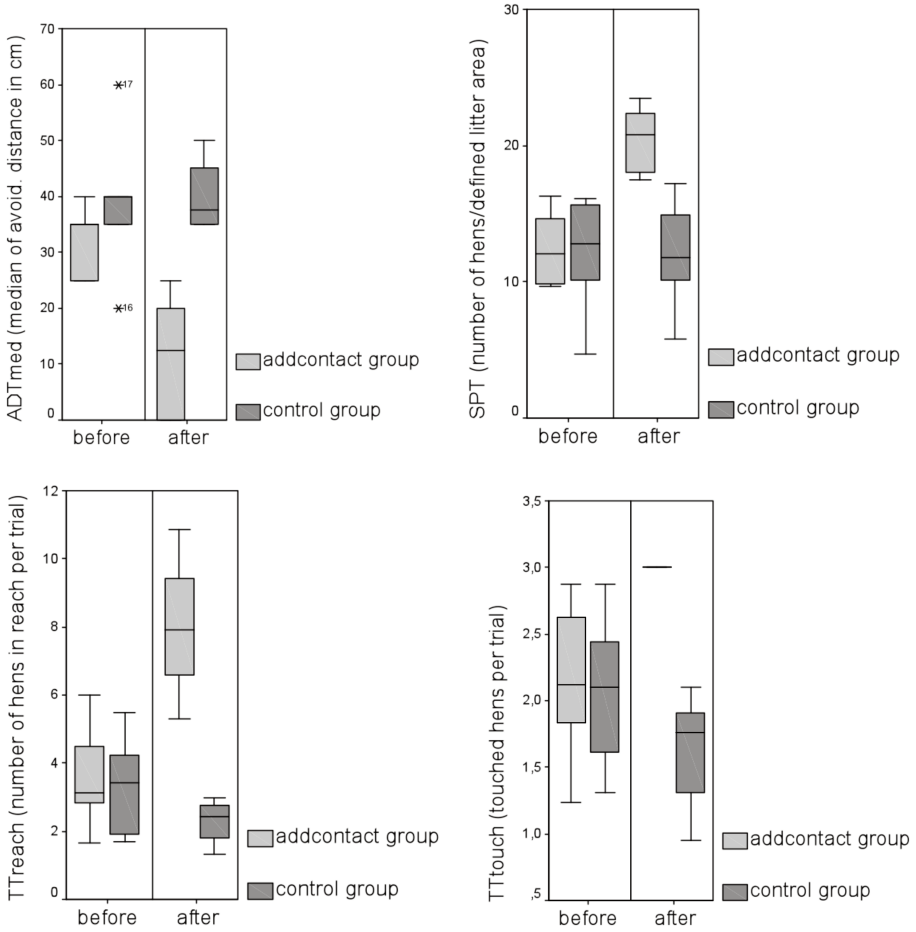


FIGURE 7.1 Differences between additional contact groups and control groups before and after a 2-week treatment period.

Source: Figures modified from Graml et al. 2008a.

Note: Birds were tested immediately before and after the treatment period with 3 tests (ADT, SPT, TTreach, TTtouch) assessing HAR.

7.5 DISCUSSION

7.5.1 VALIDITY OF THE TESTS

Validity of Tests in Non-cage Systems

The developed tests for non-cage systems turned out to be easy and fast to perform and showed a good test-retest repeatability (Raubek et al., 2007). The validity of the tests, i.e. to really measure the human-hen relationship and thus say something about the level of

TABLE 7.5 Spearman rank correlation coefficients for tests assessing HAR, general fearfulness and the standardized arena test in 10 flocks housed in non-cage systems.

					tests assessing HAR					gen. fear.	arena test	
tests	measures	median	min–max	25%–75%	SPT	TT touch	TT reach	ADT med	ADT touch	NOT	ST flyup	ST stay
SPT	Number of hens in defined litter area	5.40	.00–2.21	.00–12.88	1	.935 p=.000	.905 p=.000	–.855 p=.002	.687 p=.028	.963 p=.000	–.579 p=.079	.636 p=.048
TT touch	Average number of touched hens per trial	.71	.00–3.00	.00–2.72	.935 p=.000	1	.969 p=.000	–.945 p=.000	.757 p=.011	.972 p=.000	–.434 p=.210	.687 p=.028
TT reach	Average number of hens in reach per trial	2.74	.00–9.43	.08–6.62	.905 p=.000	.969 p=.000	1	–.884 p=.001	.651 p=.042	.954 p=.000	–.393 p=.262	.664 p=.036
ADT med	Median of distance (cm) when bird withdrew an appr. human	65.0	15–150	40–101.3	–.855 p=.002	–.945 p=.000	–.884 p=.001	1	–.815 p=.004	–.857 p=.002	.343 p=.332	–.789 p=.007
ADT touch	Percentage of hens with avoidance distance= 0cm	5.72	.00–38.10	.00–7.14	.687 p=.028	.757 p=.011	.651 p=.042	–.815 p=.004	1	.679 p=.031	–.283 p=.428	.698 p=.025
NOT	Total number of hens on defined area around novel object	3.17	.00–22.50	.23–14.78	.963 p=.000	.972 p=.000	.954 p=.000	–.857 p=.002	.679 p=.031	1	–.487 p=.154	.561 p=.092
ST flight	Proportion of birds that flew while exp. approached the table	1.67	.00–6.67	.0–4.17	–.579 p=.079	–.434 p=.210	–.393 p=.262	.343 p=.332	–.283 p=.428	–.487 p=.154	1	–.440 p=.203
ST stay	Prop. of birds that stayed in most forward area	9.0	26.67–10.0	68.33–96.67	.636 p=.048	.687 p=.028	.664 p=.036	–.789 p=.007	.698 p=.025	.561 p=.092	–.440 p=.203	1

Notes: The measure STorientation of the arena test is not described in the table because no correlations could be found with any other measure.

negative or positive emotions during interactions with humans, was shown for both tests with an on farm survey and an experimental study.

Tested on 50 flocks housed in commercial non-cage systems, tests for assessment of the human-animal relationship showed significant and moderate to high correlations with each other, while correlations with a test for general fearfulness (novel object test) were low to

TABLE 7.6 Spearman rank correlation coefficients between several measures of tests assessing HAR, general fearfulness and the standardized arena test in 14 cage housing systems (10 conventional cage systems and 4 furnished cage systems).

tests	measures	median	min–max	25%–75%	tests assessing HAR: HT	SPT	ADT med	gen. fearfulness: NOT	NOT lat	arena test: ST flight	ST stay
HT	Proportion of heads out while walking down the corridor	58.79	4.00–95.83	16.81–69.57	1	.649 p=.012	–.547 p=.043	.480 p=.083	.022 p=.940	–.414 p=.141	.507 p=.065
SPT	Proportion of heads out while standing in front of cage	16.81	1.23–5.54	11.70–25.50	.649 p=.012	1	–.788 p=.001	.530 p=.051	–.305 p=.288	–.062 p=.832	.321 p=.264
ADT med	Median of distance (cm) when bird withdrew an appr. human	3.0	15–60	23.75–6.0	–.547 p=.043	–.788 p=.001	1	–.405 p=.151	.280 p=.333	.152 p=.603	–.082 p=.780
NOT	Proportion of heads out with novel object in food trough	16.93	4.11–32.28	8.44–25.06	.480 p=.083	.530 p=.051	–.405 p=.151	1	–.538 p=.047	–.060 p=.839	–.048 p=.871
NOT lat	Latency (s) until first head out after manipulation of cage	49.15	17.60–97.60	42.08–83.68	.022 p=.940	–.305 p=.288	.280 p=.333	–.538 p=.047	1	–.502 p=.067	.409 p=.146
ST flight	Proportion of birds that flew while human approached the table	.00	.00–6.67	.00–3.33	–.572 p=.084	.114 p=.754	.039 p=.915	–.038 p=.917	–.266 p=.458	1	–.281 p=.330
ST stay	Prop. of birds that stayed in most forward area	93.33	86.67–10.0	9.0–96.67	.459 p=.182	.172 p=.365	.194 p=.591	–.121 p=.740	.292 p=.412	–.281 p=.330	1

Note: The measure STorientation of the arena test is not described in the table because no correlations could be found with any other measure.

moderate, suggesting a good specificity of the tests to measure the hens' relationship to humans (including fear of humans). The specificity of the ADT regarding the percentage of birds with avoidance distance = 0 cm (touched) was especially high (correlations to novel object test not significant with $r_s < 0.2$). Nevertheless the significant correlations of the other measures with NOT suggest that hens' reactions to a test person seem to be partially influenced as well by other motivations, e.g. neophobia, exploratory motivation or general fearfulness, as suggested before (Waiblinger et al., 2006). Those confounding factors seem to be most relevant in the SPT, correlating highest with NOT. The test

TABLE 7.7 Spearman rank-correlation coefficients of association between ages of laying hens in weeks and the results of the tests for assessing HAR or the standardized arena test in non-cage and cage systems.

Non–cage systems	Age in weeks	Cage systems	Age in weeks
SPT	–.702	SPT	–.511
	p=.024		p=.062
TT touch	–.648	HT	–.515
	p=.043		p=.059
TT reach	–.622	ADT med	.285
	p=.055		p=.324
ADT med	.683	NOT	–.300
	p=.029		p=.298
ADT touch	–.264	NOT lat	.013
	p=.461		p=.964
NOT	–.620	ST flyup	.163
	p=.056		p=.578
ST flyup	.426	ST stay	.156
	p=.219		p=.594
ST stay	–.654		
	p=.040		

procedure of SPT and NOT is very similar, both counting the number of hens approaching the object or human voluntarily. Regarding the results for the much smaller sample size of non-caged flocks tested in the calibration study we found significant and high correlations not only between the tests to assess HAR but as well with the NOT. This may be caused by (1) the use of an unknown test person which can cause neophobia especially in herds with higher general fearfulness increasing the association between the measures of HAR and NOT or (2) the procedure of the NOT involving a human filming the reaction to the NO in a distance of only 1.5 m. Thus, the NOT may at least partly measure a component of the HAR, especially in shyer herds, where hens have partly an avoidance distance close to or higher than 1.5 m. However, other procedures for the NOT are not practicable. With regard to the first point, tests using the actual caretaker as test person may have advantages, but those are not feasible due to the very variable willingness of the caretaker to test the birds and the high training efforts required for standardization.

In a further multivariate analysis regarding the 50 non-caged flocks, relationships between farm factors, stockperson attitudes/behaviour and the results of the tests were analysed. A number of factors regarding attitude and behaviour were significantly correlated with the results of the tests to assess HAR, pointing again towards validity of the tests (Niebuhr et al., 2007). This finding points as well to the conclusion that the reaction of laying hens to humans can be actively influenced by the caretaker himself, a very important point if the proposed tests are to be used in an on-farm scheme to assess animal welfare. Practical and theoretical training of stockpersons should be a means to improve HAR in laying hen flocks and minimize negative consequences.

Apart from associations to stockperson behaviour and attitudes, in further analysis ADT and TT showed significant and moderate correlations with the percentage of hens with feather damage (r_s=0.3–0.4). Stress seems to be an important factor in the development of feather pecking (El-Lethey et al., 2001) and stronger reactions of the hens to humans

during routine inspections are likely to contribute to the overall stress level. Results of the tests to assess HAR therefore seem to have a high validity to assess the overall welfare level of a given flock.

In the experimental study, the influence of two different treatments was reflected in the results of the tests in the predicted direction. Exposure to additional positive human contact led to significant differences between handling and control groups in all measures of the three tests with lower avoidance and more approach reactions in the positive handling group. The avoidance distance decreased significantly within the handling groups whereas the mean number of birds in reach and of touched birds in TT increased as well as the number of birds approaching the test person in SPT. These findings suggest that the tests effectively measure HAR and support the validity of the tests. Based on the concept of human-animal relationship reflecting previous interactions (Estep and Hetts, 1992; Waiblinger et al. 2006), additional handling is expected to alter laying hens' relationship and thus their response towards humans as shown as well in young chicks by Jones (1994, 1995) and in laying hens by Barnett et al. (1994), who used nevertheless different tests to assess HAR. For more detailed discussion of the experimental study see Graml et al., (2008a).

Validity of Tests in Cage Systems

The described tests for cage systems had to be modified for use in commercial farm with high and varying group sizes and small corridor widths. The modified tests showed significant and moderate to high correlations with each other and with another test assessing HAR, thus indicating high convergent validity. Correlations with a test for general fearfulness (novel object test) were lower and not significant (for HT $p<0.1$), indicating a good specificity of the tests, especially of ADT. Thus, a high validity of the tests could be expected.

7.5.2 Calibration between Systems

The calibration of the different measures by use of a standardised arena test (ST) failed as no significant correlations with the tests to assess HAR in the home environment were found for cage systems, although tests assessing HAR in the flock significantly correlated with the ST in non-caged systems. According to the results of the tests to assess HAR in cages we expected a high variability of results as well for the results of the arena test. Yet we found very similar results of the arena test in all flocks housed in cages.

Birds housed in cages stayed orientated forward significantly more in the ST as compared to non-caged birds. This finding suggests that the bird's reaction in the arena test depends on the housing system which unfortunately made a calibration between systems impossible. A possible explanation could be that especially in cages catching and isolation

of the birds and the new environment caused strong confounding factors disturbing the reflection of the tests to assess HAR in the results of the arena test. This problem would probably persist if a different environment for the test, e.g. a cage, was chosen, as this would mean a larger difference regarding the environment for the non-caged birds and a smaller for the caged ones.

A direct comparison of the results between non-cage and cage systems is not possible due to the different measures taken. Although the avoidance distance test produced the same type of measure (cm) in both systems, we were not able to compare them due to the different starting distances the test person used to approach the birds. Because of the small corridor width in cage system the test person was forced to approach the test birds from a maximum distance of 60 cm from the cage. In non-cage systems the test person approached the birds from 1.5 m. As we found an average avoidance distance of 70 cm in the non-cage and 35 cm in the cage systems, a shortening of the starting distance to 60 cm for both systems seems not to be advisable. It would create a lack of variance in non-cage systems and therefore make it hard to differentiate between the animals' responses in different farms. This finding might suggest that laying hens have different avoidance distances depending on the housing system. Two factors may contribute to a smaller distance in the cage systems: firstly, during regular control walks the hens may habituate to a human passing in a short distance, while in non-cage system, the regularly experienced distance to the human may be much greater for the individual hens. Secondly, the front wire mesh may be perceived as a protecting structure shortening distances. Our results suggest that the median avoidance distance was approximately half the starting distance for both housing systems. Therefore a possible solution might be to take our results into account and e.g. create a parameter like percentage of birds with an avoidance distance of 35 cm for cage and 70 cm for non-cage systems. Nevertheless this would require further investigations in a multitude of flocks especially in cage systems. For more detailed discussion of the calibration study see Graml et al. (2008b).

7.6 CONCLUSIONS

The developed tests are fast and easy to perform in both systems and seem to show a high validity.

It was not possible to calibrate the tests for non-cage and cage systems.

For non-cage systems the avoidance distance test and the touch-test seem to be the most promising tests. Both tests showed high correlations between the tests assessing human-animal relationship (HAR) and low correlations with the test for general fearfulness.

Within cage systems the ADT showed the lowest correlation with the test for general fearfulness and a high correlation with the other tests for assessing HAR. Like in non-cage systems, the SPT produced the highest correlation with the test for general fearfulness.

For practical application in an on-farm scheme to assess animal welfare we would suggest the ADT and the TT in non-cage and the ADT and HT in cage systems.

All tests described above have to be tested for repeatability which was not the aim of this study.

8

ASSESSMENT OF GENERAL FEARFULNESS

B. Forkman, T. Heiskanen, C. Graml and S. Waiblinger

8.1 SUMMARY

Tests for general fear in poultry have been reviewed in COST 846 (Forkman et al., 2007). Fear tests for poultry are among the most well researched and validated tests for fear, especially the novel object test, tonic immobility and open field tests. We suggest using the novel object test for measuring general fear in chickens at group level. The advantage is that the same type of test can be used on birds from both cage and free range systems (although in different versions).

The novel object test has been used as a fear test for birds in cages in a number of investigations and the method is well known and validated (e.g. Jones 1985a; Keer-Keer et al., 1996; Fraisse and Cockrem, 2006). The animals are tested in their home cages. The main point to observe is to test animals from different tiers since it has been shown that this has a strong effect on the level of fear (e.g. Jones, 1985a).

Loose housed animals, whether broilers or layers, will be tested in their houses. This means that all the animals that are close to the novel object are included in the test. The novel object (a brightly coloured stick) is placed on the ground; the observer starts a video-camera at the same time and walks 2 m back while filming. The filming continues for 2 minutes after which the test is terminated. Two tests are done for each shed, one just inside the shed and one in the middle of it. The use of video is however necessary if the number of birds is to be accurately assessed, especially at high concentrations. The correlation between the two tests was very good for the experience tester (Rsq=80.9%) but less so for the inexperienced one (Rsq=31.6%).

In the present study birds from 18 loose housed layer flocks (4 farms) were tested. The decoding of the video is fast and easy (3–7 minutes per test). Repeatability between decoders was very good (Spearman correlation 0.95). The repeatability between tests taken the same day, but more than an hour apart and with different observers was 0.72 (Spearman correlation).

8.2 INTRODUCTION

There are a number of different fear tests available for use on poultry, e.g. the tonic immobility test, the open field test and the novel object test. Overall the fear tests for poultry have been well validated compared to fear tests for most other farm animal species (Forkman et al., 2007). Both the tonic immobility test and the open field test require some sort of equipment, is done on one bird at a time, and is typically done on socially isolated birds. This makes both of these tests of less use for on-farm animal welfare assessment.

The novel object test is a well known and often used test for individual animals and for animals kept in cages (e.g. Fraisse and Cockrem, 2006). The advantage of the novel object test is that it uses a behavioural response that is very closely related to the definition of fear (compared to e.g. the tonic immobility test) and is therefore intuitively pleasing (Boissy, 1995). In most cases it has not been used for testing groups in cages (but see Sefton, 1976, for an exception). However, just as the novel object test for the loose housed hens is based on the individual test, we can use the test for the individually caged birds to test the group caged birds.

8.3 METHODS

Because of the problems mentioned in the introduction all the tests were done in Austria. Eighteen different flocks of loose housed layers were visited. These came from four different farms with from four different farms with 2, 2, 2 and 12 flocks respectively. The persons doing the tests were Christine Graml who has extensive experience in doing the test and Teresia Heiskanen from Denmark who had not previously done the test. The birds were between 19 and 53 weeks old and all but one flock were Lohmann Brown (the exception was one flock of ISA Brown). The group size was between 400 and 3000 birds.

The test done was a novel object test, as the novel object a 30 cm long stick with bands of bright colours was used.

To standardise the test the zoom of the camera used was adjusted so that at 2 m distance it covered 100 by 130 cm.

Each test started with one person entering the shed and standing still for five minutes just inside the door, this to get the hens a chance to discover the test person and to calm down. The test person then slowly walked 2 meters into the shed and put the novel object onto the floor. While holding the hand on the novel object the person then turned on the camera and while filming the novel object walked 2 meters back. The camera was always focused

on the middle of the novel object. Filming continued for 2 minutes and 10 seconds. The novel object was retrieved and the test was repeated approximately halfway into the shed (but now without the initial 5 minutes standing).

The test sequence was repeated by the other test person, more than one hour later. The order of the test persons was semi-random.

8.4 RESULTS

For the experienced test person there was a very good correlation between the tests done just inside the shed and in the middle of the shed (Rsq=80.9%, N=17, P= 0.000). For the inexperienced person this correlation was less (Rsq=31.6%, N=18, P=0.021).

The flocks were ranked on the level of fearfulness shown as the sum of number of birds inside the 20 cm frame, and if two flocks had the same number, total number of birds inside the video frame. The rank of the farms assigned by the experienced and inexperienced tester, based on the mean level of fearfulness (test 1 and 2), was compared. Despite the tests being done by two different test persons and with at least one hour in between, there was still an acceptable level of correlation: 0.72 (N=17, Spearman correlation, one missing value).

8.5 CONCLUSIONS

The novel object test seems to be a feasible and repeatable test for fear in loose housed poultry. The literature available indicates that the same holds true for battery caged hens.

9

MEASURING DISTRESS VOCALISATION IN POULTRY

B. Algers, L. Nordensten and P. Zimmermann

9.1 SUMMARY

Layer chickens and broilers both have specific vocalisations that can be regarded as emotional expressions of distinct states of the animal. These vocalisations can be used as indicators of the welfare of the animal. For example, laying hens that are thwarted in the performance of behaviour produce a specific vocalisation, the gakel-call, which signals frustration (Bäumer, 1962; Schenk et al., 1989; Zimmerman and Koene, 1998). Broilers have a distress call that is given, for instance, when they experience fear or pain (Wood-Gush, 1971; Collias, 1987).

The aim of this study was to determine whether vocalisations can be used as a non-invasive, animal-based method to assess poultry welfare on a group-level, on-farm and at pre-slaughter handling. For this purpose, we investigated whether it is possible to devise a simple, reliable and repeatable method of recording poultry vocalisations on a group-level.

An experimental study was carried out in which the relation between frustration and vocalisations in 6 groups of 4 adult laying hens was studied. These hens were subjected to three different treatments: deprivation of food, deprivation of nest, and no deprivation (control treatment). Qualitative analysis of vocalisations showed that during deprivation more gakel-calls were given than during non-deprivation, indicating that during deprivation hens experienced frustration. The hens gave more gakel-calls when they were deprived of food than when they were deprived of a nest. Quantitative analysis of vocalisations showed that different vocalisations, as identified by the human ear, had distinct temporal and spectral patterns. However, analysis of vocalisations was hampered by background noise caused by environmental sounds such as fans and sounds made by other hens.

Furthermore, a method for the assessment of distress vocalizations (i.e., high-pitched vocalisations or HPV's) in abattoirs and in on-farm situations was tested and evaluated. In the abattoir, HPV's were recorded at the point of shackling, where the animal generally

experience stress as a result of handling prior to stunning. Both in the abattoir and the on-farm situation, we used a combination of one-zero and instantaneous sampling to record HPV's. Our tests showed that it is very difficult to accurately discriminate between HPV's on the one hand and other chicken sounds and environmental noise (e.g. fans, feeder chain) on the other hand.

The general conclusion is that our method of monitoring frustration-calls and distress-calls in groups of domestic chickens is not a promising one; accurate discrimination of these calls is seriously hampered by sounds from other chickens and environmental noise.

9.2 INTRODUCTION

Layer chickens and broilers both have specific vocalisations that can be regarded as emotional expressions of distinct states of the animal. These states can be used as indicators of the welfare of the animal. For example, laying hens that are thwarted in the performance of behaviour produce a specific vocalisation, the gakel-call that signals frustration (Bäumer, 1962; Schenk et al., 1989; Zimmerman and Koene, 1998). Broilers have a distress call that is given, for instance, when they experience fear or pain (Wood-Gush, 1971; Collias, 1987).

The purpose was to develop a non-invasive, animal-based method to assess welfare on group level. The aim was to study how specific vocalisations can be used to assess welfare in poultry, to quantify the repeatability and reliability of a simple method for recording the frequency of HPV's (high pitched vocalisations) in groups of layers/broilers by an observer during the time when they are being inspected by the keeper. Shackling and slaughter of poultry has long been an area of great concern from a welfare point of view (Weeks and Butterworth, 2004). A correct assessment of High Pitch Vocalisations, HPV's, is of great concern to assure a high standard of poultry welfare. Therefore, these vocalisations can be used to monitor the welfare status of layers and broilers, either on an individual or a group level. An area where vocalisations could prove to be a useful measure of instantaneous welfare is in the pre-stun handling were procedures cause pain and induce stress in poultry (Raj, 1998; Weeks and Butterworth, 2004). Vocalisations seem to be strongly associated with stressful events at pre-slaughter handling in poultry as uncrating and shackling of live birds.

Slaughter lines in Europe today commonly operate at speeds of 100 to 140 birds per minute (Bilgili, 1999). There are reasons to believe that some proposed signs of impaired welfare will be very difficult to detect during the regular slaughter process due to high speed of the shackle line. Therefore, a different way of monitoring, as with vocalisations which would exclude the need to monitor each bird individually, would be welcomed. The aim of this study was to develop and evaluate a method for clinical assessment of distress

vocalisations in poultry during shackling and on-farm situations, possible to conduct under commercial conditions.

Furthermore, distress-calls and gakel-calls were studied under experimental conditions in which birds were deprived of important resources, such as food and nest, to investigate the relation between frustration/stress and vocalisations.

9.3 MATERIAL AND METHODS, FIELD OBSERVATIONS

9.3.1 During Shackling

Visits were made to three European abattoirs. Two abattoirs conducted electrical stunning and the remaining one conducted CO_2 + Ar gas stunning. A water bath electrical stunner is the most common method used to slaughter poultry under commercial conditions (Raj, 1998; Bilgili, 1999). After test visits to the slaughter plants, a method for assessment of HPV was evaluated. At the abattoir, monitoring of vocalisations was first conducted at three different locations;1) the shackle area 2) the stunner entrance and 3) the transport unloading area. Vocalisations were recorded at all locations. While placed in the transport unloading area, the birds were relatively calm and no HPV was recorded by the observer. At the stunner entrance, the noise pollution from the birds being shackled was confounding the assessment. After evaluation it was decided to only monitor at the shackle station.

The birds were monitored for HPV while being shackled during a 25 min period with intervals of 20 seconds. Visual monitoring of 3 shackling stations was conducted simultaneously to correlate occurring vocalisation to shackled birds. Recordings were made using prepared protocols with rows as intervals of 20 sec. During each 20 sec interval, the observer listened whether any of the birds shackled vocalised or not (one-zero sampling). At the tone, marking the end of the interval, the observer noted whether any of the birds were vocalising at that very moment (instantaneous sampling). Additionally, in a third column the observer noted if it was a single bird vocalising or several birds, thus modifying the information gained in the previous column.

Three observers were trained by being instructed how to monitor the defined behaviour at the final stages of driving and listening to recorded examples of the HPV. After training, monitoring from a video clip was done simultaneously by the observers. It was not possible to test inter-observer reliability under commercial conditions, due to industry being reluctant to have observers at their premises because of Avian Influenza.

Statistical analysis was performed in Minitab (MINITAB® Release 14.1, Minitab Inc, 2003). The combined monitoring results of all three observers were statistically analysed for validation and inter-observer reliability. The inter-observer agreement was evaluated

by analysis of Fleiss' kappa test (Fleiss, 1971). 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 (1991) 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.

Fleiss' kappa coefficients (Fleiss, 1971) for inter-observer reliability during monitoring of driving into butina/stun box were calculated by using the 'Attribute Agreement Analysis' function in Minitab (MINITAB® Release 14.1, Minitab Inc, 2003).

9.3.2 On-farm Flocks – Layers

Five separate layer farms were visited a total of nine times each, in order to detect possible confounding factors. The layer farms visited had flocks of 200, 1200, 3000+2900, 5000 and 30 000 individuals, different strains and ages that ranged from 20–72 weeks. The housing systems were mainly Redell or Vencomatic system with Jansen nests.

Visits were made during the morning inspection, just before and during egg collection. The lights went on between 2 am and 7 am. The morning inspections usually started 3–4 hours after lights-on. The monitoring was conducted following the keeper's pace and movement while inspecting. This was to reduce the observer effect, as hens are likely to be affected by novel persons (ref.). Factors that may influence poultry behaviour, such as flock size, animal/area ratio, age, strain, artificial light intensity, artificial manipulated circadian pattern, temperature, interior design, background noise, keeper-induced events, food access etc., were noted down on a prepared sheet and taken into account during evaluation.

No statistic evaluation of the method was carried out.

9.3.3 On-farm Flocks – Broilers

Two broiler farms were visited on one occasion each. In order to detect possible confounding factors, flock ages differed between the farms (29 and 32 days old and 11 and 14 days old). Housing system and broiler strain are kept the same in the country where the field visits were made. The circadian pattern, temperature and light intensity in the stable were artificially kept correlated with the flock age.

Visits were made during the morning inspection. The monitoring was conducted following the keeper's pace and movement while inspecting.

No statistic evaluation of the method was carried out.

9.3.4 Experimental Flocks – Layers

For infectious disease control reasons (Avian Influenza), the experimental treatments were carried out at the SLU poultry research facility, Skara, Sweden in June 2006, instead of on a commercial farm. The focal animals (layers) were housed in six pens in groups of four individuals. The habituation period was 4 days. Groups were subjected either to deprivation of food, deprivation of nest or no deprivation. The treatments were carried out and behaviour and vocalisations of the groups were filmed during three consecutive days. Three video cameras were used and the six groups were divided into two subgroups that were filmed either in the morning or in the afternoon. Broilers were not used in the experimental treatments due to severe noise in the experimental room, caused by a large group of cockerels in the adjacent room of the experimental poultry research facility, Skara.

All animals in the set up were used to human presence and handling. The groups were of different age which simulated the difference in flock age between different farms.

In the food deprivation treatments, the feeder was removed from the pens at 2 pm the day before the first observation day, and at 2 pm after recordings had been carried out on the following days. In the nest deprivation treatment, the nest was removed at 2 pm the day before the first observation day, and at 2 pm after observations had been carried out on the following days. The control groups had free access to food and nest at all times. All groups had had access to water at all times. The temperature was kept at 20°C. At 2 pm, after the observations were done, water was changed and the feeder was refilled. Then eggs were collected and counted. All necessary handling of the layers took place at this point, to minimize the possible effect on the treatment observation on the following day.

Each observation consisted of two 30-minute periods (see also Table 11.2): in the first period, the deprivation-period, hens were still deprived of food/nest or not (control treatment). In the second period, the post-deprivation period, feeders and nests were returned to the pen. It was anticipated, due to the limited space around feeders and in the nest, that competition would arise and possibly frustration as a result. In the control treatment, the experimenter entered the pen and left it again to control for possible experimenter effects on behaviour and vocalisations in the post-deprivation period.

TABLE 11.1 Overview of groups and individuals in the experimental flock of layers.

Group	Individuals
Group 1	755, 441, 447, 475
Group 2	156, 712, 728, 801
Group 3	725, 811, 826, Leg X
Group 4	463, 753, 816, Leg X
Group 5	1504, 1538, 1549, 1628
Group 6	218, 241, 266, 227

Notes: n16 × L13 Leghorn strain: groups 1–4, batch December 2005; n4 × vitiligo strain: group 5, batch 23 October 2005; n4× F1 OS strain: group 6, batch June and July 2005.

TABLE 11.2 The groups and the treatments.

	Observation day 1	Observation day 2	Observation day 3
Group 1	E no food	L no nest	E non-dep
Group 2	L non-dep	E no nest	L no food
Group 3	E no nest	L no food	E non-dep
Group 4	L no food	E non-dep	L no nest
Group 5	E no food	L no nest	E non-dep
Group 6	L non-dep	E no nest	L no food

Notes: no food: food deprivation; no nest: nest deprivation; non-dep: control situation. E and L indicate early observation and late observation, respectively.

Statistical analysis was carried out in SAS using an ANOVA with treatment, group and period as factors, and correcting for repeated measurements within groups. Numbers were log-transformed if necessary.

9.4 RESULTS

9.4.1 During Shackling

The inter-observer Kappa value was found to be low (for a first trial with naive observers) for monitoring chickens being shackled (0.24, instantaneous sampling, or a 47% match). The one-zero sampling assessed at the same trial had a Fleiss' kappa value –0.06. The negative value is due to a lack of variation in HPV calling rate in the observed material (100% HPV through all observed intervals, i.e. leaving only the possibility to score 1, instead of 1 *and* 0), making assessed samples identical.

9.4.2 Experimental Flocks – Layers

We found an effect of treatment on the number of gakel-calls ($F_{2,16}$=6.21, P=0.01). More gakel-calls were given in the deprivation treatments (food: 0.17±0.02, nest: 0.13±0.02; mean number per minute±SEM) than in the control treatments (0.04±0.01). There was no significant differences between treatments for the HPV's (food: 0.00±0.00, nest: 0.01±0.00, control: 0.00±0.00).

We found an effect of treatment on the number of gakel-calls in the deprivation-period ($F_{2,16}$=3.08, P=0.08). During food deprivation birds tended to produce more gakel-calls than during nest deprivation (food: 0.19±0.05, nest: 0.13±0.03).

We found an effect of period on the number of agonistic interactions ($F_{1,13}$=7.13, P=0.01) and gakel-calls ($F_{1,13}$=3.17, P=0.09). More aggression was shown during the post-

deprivation (0.17±0.03) period than during the deprivation-period (0.11±0.01). Birds tended to give more gakel-calls during the post-deprivation period (0.07±0.01) than during the deprivation-period (0.11±0.02). We found a correlation coefficient of 0.32 between number of agonistic interactions and gakel-calls.

We did not find an interaction effect of treatment and period for any of the vocalisations.

9.5 DISCUSSION

The evaluated assessment method of monitoring HPV's in an on-farm situation and in abattoirs is animal-based and objective and aimed at reducing subjective opinions between observers. This sampling method provides a crude measure, but we are uncertain whether it will be possible to obtain a higher resolution than we obtained in our study. What has been assessed during this study is in fact the difference in vocalisation levels between two different type of abattoirs (electrical stunning and gas stunning) and not between the same type of abattoir.

Abattoirs: To measure vocalisations in an abattoir with a high-speed slaughter line proved to be hard due to high noise level from fans and vocalisations from other shackled birds. The high speed of today's slaughter lines makes monitoring of several important behavioural signs under commercial conditions very difficult. To find a way around the difficulties and film the shackle area for a later analysis back at the office proved impractical as it was impossible to determine which individual produced the vocalisation from film footage. The test persons found it hard to analyse vocalisations from video clips. As the test leader had the stopwatch, to make sure the signal marking instantaneous sampling came simultaneously for all test persons, they felt startled as the signal came, not being able to keep up with the stopwatch. They expressed observer fatigue and lack of focus between signals. If training should be conducted it should be done live at the abattoir.

On-farm – Layers: The HPV's heard in layer flocks upon inspection of the flock seem to be very similar to the alarm-call that indicates a ground predator illustrated by Gyger et al. (1987). This indicates that it is actually the instantaneous stress induced by the inspection itself that is measured rather then the overall welfare state of the animals.

However, an objective monitoring system would be much welcomed by the producers. The producers involved in this study did express concerns about experiences of arbitrariness in previous animal-welfare inspections.

On-farm – Broilers: In poultry, the apprehension of man is instinctive and present from hatching (Fraser and Broom, 1990). Poultry are neophobic and unknown people, as inspectors, may cause fear and evoke flight behaviour. Broilers were alarmed at inspection

and emitted distress-calls and it is not possible at that age to achieve a complete habituation to man. However, it was impossible to relate distress-calls to specific behaviours, such as aggression, because of the large number of birds per flock.

Experimental flocks – Layers: We found a relationship between frustration and the number of gakel-calls as been previously found in other experimental studies (Zimmerman and Koene, 1998, Zimmerman et al., 2000). Monitoring of different vocalisations was facilitated by recording behaviour and vocalisations on tape. However, it was still difficult to identify which individual was giving the call, due to sounds from other hens and environmental noise in the experimental room.

9.6 CONCLUSIONS

Vocalisations are generally considered to be a good welfare indicator in poultry (Zimmerman and Koene, 1998 and other refs). However, this mainly applies to experimental conditions in which small groups of hens are studied. The evaluated method is a poor candidate for monitoring poultry vocalisations in poultry under commercial conditions because of serious problems with accurate identification of vocalisations.

10

POSITIVE EMOTIONS IN POULTRY

L.J. Keeling, and P. Zimmerman

10.1 SUMMARY

Animal welfare has been traditionally defined as the absence of suffering and negative emotions like fear and frustration. If pleasure is mentioned at all, it is seen as of less importance. However, because the emphasis is on avoiding suffering, little is known about positive emotions; what our farm animals actually perceive as pleasurable or how they act under pleasurable circumstances (Boissy et al., 2007).

It has been proposed that the feedback to animals is positive (both in terms of increased motivation to perform the behaviour and the mental experience) when performing the right behaviour at the right time. Feeding is positive to a hungry animal and, if the food is especially tasty, it can be positive even to a non hungry animal. In the first experiment we studied positive anticipation behaviour in poultry to find out whether anticipating something positive (food) was associated with specific behaviours which could then be used as indicators of positive emotional states. This was done by training adult laying hens to associate a particular sound cue with either the opportunity to get access to a food reward (positive situation), to be squirted with water (negative situation) or to nothing happening (neutral). Behaviour and vocalisations were recorded during the 15-sec anticipatory period that the sound cue was presented. As a baseline measurement, behaviour and vocalisations were also recorded in a 15-sec period before playback of the sound cue. We found a significant increase in the performance of standing alert and comfort behaviour (wing-flapping, feather ruffling and scratching) in anticipation of a reward. However, these are not changes in behaviour that can be readily observed under commercial conditions during a short visit.

In the second experiment we studied play behaviour in relation to housing conditions in broilers in three different housing conditions (barren, normal and enriched). Play behaviour is most frequent in young animals and occurs mainly when all other essential needs are satisfied. Thus it probably indicates that necessary resources are present and, at least at for that moment in time, the animal is not experiencing any negative affective states. We found that total play behaviour decreased over time in all treatments and, contrary to our predictions, that there was overall least play in the enriched pens. The introduction of play objects was successful in eliciting play behaviour and in all pens more time was spent on

play behaviour (play-fighting, play-running and play-running+wing-flapping) when the play objects were present. However, again contrary to predicted, we found that chicks in normal pens performed more play behaviour when the play objects were present than chicks in enriched pens and that play behaviour in the barren pens did not differ from either the normal and enriched pens. It was concluded that with our current knowledge of play behaviour in birds, it could not be used as a valid indicator of positive emotions in poultry and that more research was needed.

10.2 POSITIVE ANTICIPATION

10.2.1 Introduction

It is proposed that behavioural indicators of positive emotional states can be identified by experimentally inducing such as state and observing the behaviour. One of the approached to induce such a state in animals is the anticipation or expectation of a positive event or stimulus (Spruijt et al., 2001; Boissy et al., 2007). Positive anticipation can be defined as 'responses elicited by rewarding stimuli that lead to and facilitate consummatory behaviour' and it is claimed to be rewarding in itself through the conditioned release of opioids, allowing it to replace consummatory acts (Spruijt et al., 2001). The types of behaviour that are displayed during anticipation of a reward differ per species. In rats, anticipation is reported to be characterised by increased activity and behavioural transitions (Van der Harst et al.. 2005, Von Frijtag et al., 2000) and an increased frequency of the occurrence of exploratory behavioural patterns (Spruijt et al., 2001). Minks are also reported to increase their amount of behavioural transitions (Vinke et al.. 2004), but cats are known to decrease these transitions during anticipation (Van den Bos et al.. 2003).

In the present study, anticipatory behaviour was studied in poultry. Laying hens are known to pay attention to complex combinations of cues and produce finely tuned responses in response to them. For example, they are known to be able to estimate time, as domestic hens have shown to be able to maximise their pecking response 6-8 minutes after being given signal indicating that pecking will be rewarded after a wait of 6 minutes (Abeyesinghe et al., 2005). The combination of paying close attention to environmental cues and being able to estimate time makes them suitable candidates for anticipation research.

This first study aimed to identify indicators of positive welfare through reward anticipation. We trained laying hens to associate a specific sound cue with either a positive, a negative or a neutral event and we observed the behaviour shown by the birds.

10.2.2 Methods

For this study 24 adult LSL hens were trained, of which 18 were used in the final experiment. They were divided into 6 different groups, of 3 birds each. Each group was given a unique combination of sound cues and orientation of the test pen (see Table 10.1).

At the start of the training phase the hens were housed in pens measuring 150 x 150 cm in groups of 2 to 5 birds per pen. They had ad libitum access to food (standard layer crumble) and water in the home pens, except in the first eleven days of the training phase when they were deprived of food for 1 to 5 hours prior to training. The floor of the pens was covered in wood shavings and the birds had access to 2 perches over the full width of the pen and one nest box. Lights were on from 8.00 a.m. until 20.00 p.m.

The experimental pen (see Figure 10.1) measured 200 x 125 cm and was 60 cm high. It consisted of two compartments of approximately equal size (100 x 125 cm). The two compartments were divided by wire mesh and connected by a push-door which could be locked by an electromagnet.

TABLE 10.1 The experimental groups.

Experimental group	Orientation	Positive cue	Negative cue	Neutral cue
1a	Normal	Beep	Music	Ring
1b	Inverted			
2a	Normal	Music	Ring	Beep
2b	Inverted			
3a	Normal	Ring	Beep	Music
3b	Inverted			

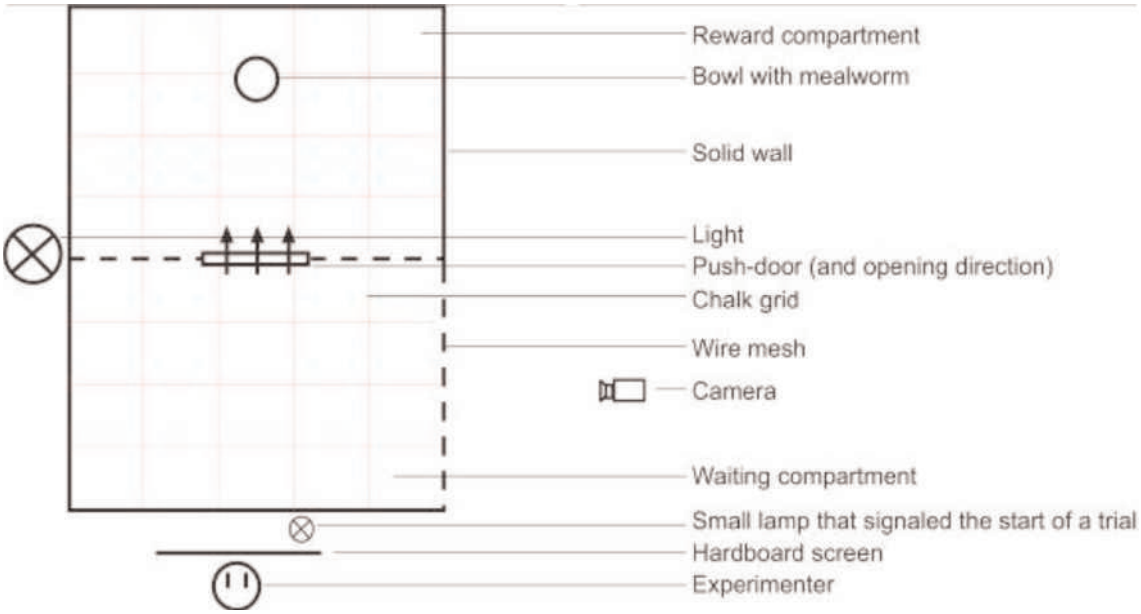


FIGURE 10.1 Drawing of the experimental pen.

The first compartment, or 'waiting compartment', had a wire-mesh wall on one side to allow video recordings to be made from an almost horizontal angle.

During training the hens were first habituated to the experimental pen and trained to go from the waiting compartment to the reward compartment through the push-door. In the reward compartment they received a reward of 2 mealworms in a blue bowl. In the next phase of training the hens were trained to associate this positive event of access to 2 mealworms with a sound cue that lasted 15 seconds. After they reached criterion in this training phase they were also trained with the negative (a squirt of water from a water pistol) and neutral events (nothing happened), each with their own sound cue for each experimental group.

The criterion that was used to assess whether a birds had learned the procedure was whether it approached and pushed the door within 5 seconds after the light was switched on, and did not push it at other times. Eighteen animals were selected for use in the experiment (randomly within each test group).

During testing all birds were tested every day for 4 consecutive days in a row, in one session a day. One session consisted of eight trials per bird (5 positive (POS) trials, 1 negative (NEG), 1 neutral-with-sound (SN) and 1 neutral-without-sound (MN) trial). These trials were ordered in such a way that a session always started with a positive trial, and in the rest of the session a positive trial was equally often preceded by a POS, NEG, SN or MN trial. Each trial consisted of a waiting period of 15 secs, followed by a 15-sec anticipation period in which the sound was played (POS, NEG, SN) or nothing happened (MN).

10.2.3 Results

In the waiting period the hens, who were supposed to be unable to predict the treatment in the succeeding anticipation period, did not show differences for most of the behaviours between treatments. However, the incidence of gakel-calls in the waiting period was higher in the SN treatment ($P<0.05$) compared to the NEG and MN treatments (SN=0.36±0.12, NEG=0.23±0.11, MN=0.21±0.08).

In the POS, NEG and SN treatment, chickens showed higher levels of head-movements ($P<0.001$) and head-flicks ($P<0.001$) and a lower incidence of ground-pecking ($P<0.001$) in the anticipation period compared to the preceding waiting period (see table below for means±SEM). In the POS treatment the chickens gave fewer gakel-calls ($P<0.05$) and performed more standing alert ($P<0.05$) in the anticipation period than in the waiting period. In the SN and NEG treatment the chickens showed an increase in locomotion ($P<0.05$) in the anticipation period compared to the waiting period.

When comparing the anticipation periods of all treatments we found that incidences of standing alert were higher ($P<0.05$) in the POS treatment (0.25±0.05) compared to the NEG treatment (0.11±0.05). Also comfort behaviour (wingflapping, ruffling and

TABLE 10.2 Mean incidence of behaviours in waiting period and the succeeding anticipation period.

Treatment	Behaviour (incidence)	Waiting period	Anticipation period
POS	Head-movement	17.98±.50	23.77±.92
	Head-flick	.03±.00	1.35±.36
	Ground-pecking	2.01±.38	.43±.07
	Gakel-calls	.29±.08	.11±.04
	Standing alert	.13±.03	.25±.05
NEG	Head-movement	17.93±.59	27.72±1.32
	Head-flick	.05±.03	1.48±.43
	Ground-pecking	1.40±.29	.07±.05
	Steps	2.62±.25	8.87±.96
SN	Head-movement	19.44±.74	24.94±1.03
	Head-flick	.0±.0	1.54±.37
	Ground-pecking	1.81±.47	.25±.10
	Steps	3.54±.41	5.65±.50

scratching) occurred more ($P<0.05$) in the POS treatment (0.05±0.02) compared to the other treatments (NEG=0.0±0.0, SN=0.01±0.01, MN=0.01±0.01).

10.2.4 Discussion

Laying hens were found to be able to discern between different sound cues that predicted different events. In anticipation of a positive event, birds showed significantly more comfort behaviour and more standing alert compared to a neutral event. In anticipation of a negative event they showed more head movements and locomotion compared to a neutral event. These results seem to indicate a general relaxed state during positive anticipation, and a general aroused state during negative anticipation.

It was regarded that changes in the frequency of standing alert and comfort behaviour had little potential as feasible measure of positive emotions in laying hens under commercial conditions so we did not proceed with our planned study to validate this measure using pharmacological manipulations.

10.3 PLAY BEHAVIOUR

10.3.1 Introduction

Play is known to be reduced in adverse physical and environmental conditions and following severe or prolonged negative emotions (Spinka, 2001; Donaldson et al., 2002). A reduction in play behaviour may indicate endangered health or welfare (Lawrence, 1987;

Newberry et al., 1988). On the other hand, play provides the individual with valuable experience that enhances later survival. Since these are long-term benefits, it is proposed that a proximate reward mechanism (e.g. a positive emotional state) must be involved to ensure that animals do indeed engage in play at appropriate times, rather than performing some other adaptive behaviour, such as resting and conserving energy. Pigs show a reduction in play and explorative behaviour in barren housing conditions compared to enriched housing (Bolhuis et al., 2005), but little is known about the interaction between play in chickens and their housing environment.

This experiment aimed at studying (the development) of play behaviour under three different housing conditions, Enriched, Barren and Normal. It was predicted that in Enriched housing the conditions to perform play and species-specific behaviours would be maximised. The total amount of play was expected to be least in the Barren environment. We also tested whether providing broiler chicks with (extra) enrichment ('toys') evoked play behaviour and also how the level of this evoked play behaviour differed between the three housing conditions. The prediction was that chicks from barren conditions would show more play behaviour than chicks from the other two housing conditions.

10.3.2 Methods

There were 20 pens (2 x 1.45 m) of 8 broilers each. We had 4 Barren (no wood-shavings, no perches) pens, 8 Normal pens (wood-shavings) and 8 Enriched pens (wood-shavings+scattered whole-oat every 2–3 days, perches at 10 and 30 cm height).

Pens were observed for 10 minutes a week, during which all occurrences of presumed play behaviour were recorded (Kruijt, 1964). One focal animal per pen was observed for 10 minutes a week using instantaneous scan sampling with an interval of 10 seconds. Furthermore, each pen was filmed for 30 minutes, once a week. Two focal animals per pen were observed from these video recordings, using focal continuous sampling.

At 2 weeks of age half of the pens in each treatment were provided with toys (small plastic toothpicks, a ball, a cardboard box) during 30 minutes. This period coincided with the filming of each pen. At 4 weeks of ages, the other half of the pens got the toys. At 6 weeks all pens were given the toys.

10.3.3 Results

The live observations showed an effect of week on the incidence of play-running ($P<0.01$), play-running+wing-flapping ($P<0.001$) and a trend for wing-flapping ($p=0.06$) and total play behaviour ($P=0.07$). We found an interaction effect of week*housing for play-running+wing-flapping ($P<0.01$). Chicks showed more bouts of play-running in week 1 (mean±sem: 3.72±1.13) compared to weeks 5 (0.67±0.19) and 6 (0.67±0.18) and more play-running+wing-flapping in week 1 (3.30±1.08) compared to all other weeks (1.32±0.18). In week 1, chicks from the Barren pens showed more play-running+wing-

flapping (P=0.05) than chicks from the Normal pens (Barren: 8.00±4.16, Normal: 0.93±0.71). The Barren chicks showed a decrease in play behaviour from week 1 onwards; the Normal and Enriched chicks did not show such a decrease.

The video observations showed that without toys present the Enriched pens tended to spend a smaller percentage of time on play behaviour (P=0.077) than the Normal and Barren pens (see Figure 10.2).

The Enriched pens also tended to show less spar-fights (P=0.09) than the Normal pens (Enriched: 0.23±0.09, Normal: 1.50±0.35).

The video observations of weeks 1 to 5, without toys, showed that birds tended to show (P<0.10) more bouts of play behaviour in week 1 compared to week 5 (Week1: 6.5±1.4, Week5: 4.3±1.5). Birds spent more time on play-running (P<0.05) in week 1 than in week 5 (Week1: 0.67±0.14, Week5: 0.39±0.19, expressed as percentage of observation time). However, time spent on spar-fighting tended to be higher (P<0.10) in weeks 4 and 5 compared to week 1 (Week1: and the incidence of wing-flapping tended to be higher in weeks 2 and 3 compared to week 5.

The analysis of video data also revealed that the introduction of toys elicited play behaviour. In week 2 a larger percentage of time (P<0.05) was spent playing with toys present (3.01±0.94) than without toys present (0.58±0.22). In week 4 the chicks also spent more time playing (P<0.05) when the toys were present (4.63±1.93) than without toys (1.12±0.34, see Figure 10.3).

With toys present the chicks in the Normal pens spent a higher percentage of time playing (P<0.05) than the chicks in the Enriched pens; the Normal chicks spent more time pecking

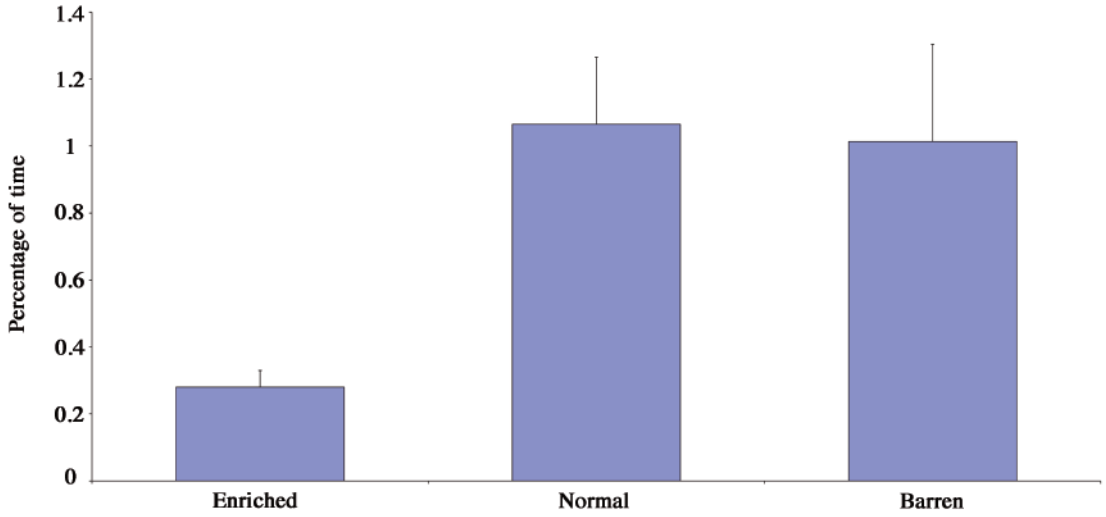


FIGURE 10.2 The mean percentage of time spent on play behaviour in the video observations without toys present.

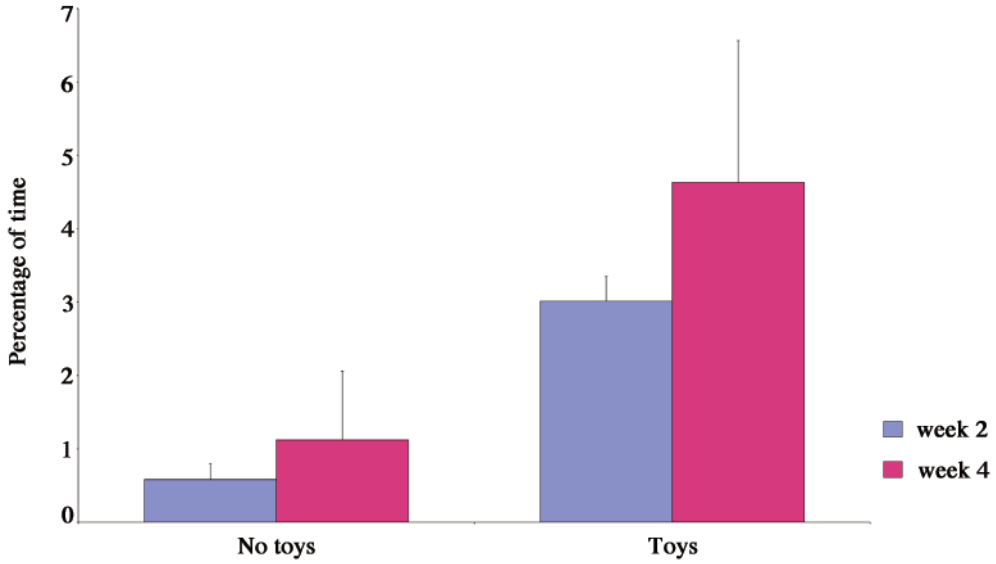


FIGURE 10.3 The mean percentage of time spent playing in all housing condition without and with toys in weeks 2 and 4.

(0.05) at the toys, they showed more bouts of play-running ($P<0.001$) and showed more wing-flapping ($P<0.05$).

10.3.4 Discussion

The live observations of the play behaviour of broiler chicks during 6 weeks only showed a difference in week 1 between the housing conditions: Barren chicks showed more play-running+wing-flapping than Normal chicks. We found a decrease in play behaviour in Barren pens, but not in the Normal and Enriched pens.

The video observations showed that Enriched chicks spent less time on play behaviour than the Normal and Barren chicks when no toys were present. This is in contrast with our hypothesis that chicks in enriched conditions would show higher basal levels of play behaviour compared to normal and, especially, barren conditions.

The video observations showed that the toys were successful in eliciting play behaviour during the 30 minutes they were present in all housing conditions. When the toys were present, the Normal pens showed more play behaviour than the Enriched pens. This is in line with findings in pigs who spend more time interacting with novel objects when housed under barren conditions compared to enriched conditions (Bolhuis et al., 2005). Why the Barren chicks in our study did not show the highest level of play behaviour with toys present is not clear. It could be that because of the barren conditions the chicks became less responsive to the novel toys. Since the occurrence of play behaviour was not in accordance with the predictions it was decided that at present it could not be used as an indicator of positive emotions in poultry

TABLE 10.3 Means of the Normal and Enriched pens from the observations with toys present.

	Normal pens	Enriched pens
Time spent on play	6.62±1.90	2.58±0.93
Time spent on pecking at toys	5.48±1.63	2.50±0.94
Time spent on locomotion	16.14±2.34	9.96±0.05
Bouts of play-running	6.33±2.50	0.75±0.25
Number of wing-flaps	3.93±0.86	1.37±0.39

10.4 CONCLUSIONS

Neither of the two studies reported here identified behaviour patterns that could, with our present state of knowledge, be regarded as valid indicators of positive emotions in poultry. In the first study differences in behaviour were found, but the indicators – an increase in standing alert and comfort behaviour – were too general to be considered reliable indicators in commercial practice. Further it is not clear what the baseline level should be under commercial condition and so, from what level there should be an increase. In the second study the occurrence of play behaviour was not according to the predictions, leading us to question whether the behaviour we recorded was true play behaviour or, alternatively, that play is not a valid indicator of positive emotions. We still feel that there is potential to identify indicators of positive emotional states in poultry. But it is clear that more work is needed with anticipation and play behaviour and perhaps new avenues need to be explored.

11

QUALITATIVE BEHAVIOUR ASSESSMENT

F. Wemelsfelder, U. Knierim, H. Schulze Westerath, T. Lentfer, M. Staack, and V. Sandilands

11.1 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, sheep 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 asks observers to develop their own descriptive terminologies, and 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 poultry expression.

On the basis of previous QBA research with poultry and consultation with poultry experts, we designed a rating scale of 28 descriptors. This scale was tested by three assessors in 23 laying hen houses on 17 farms in different regions of Germany. Assessors were given detailed instructions on the procedures of assessment and the use of the rating scale. The inter-observer reliability of the QBA scores attributed to the 23 hen units was tested using Kendall Correlation Coefficient W, and showed satisfactory reliability ($W \geq 0.70$) for 20 out of 28 descriptors. Comparison of Principal Component Analyses (PCA) of each assessor's scores showed similar emergent patterns of poultry 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 8 descriptors with low loadings and low apparent welfare relevance from the assessor data sets. PC1 of the 'reduced' PCAs explained 58–75% of the total variation between flocks (depending on the assessor), and showed good inter-observer reliability (Kendall $W = 0.83$; $p<0.001$). PC1 scores also showed a high

mean correlation (0.88; $p<0.001$) to the assessors' scores of the descriptor 'welfare overall', which supports their direct relevance to poultry welfare.

Thus we propose that PC1 may provide an integrative measure of positive and negative poultry 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 poultry, and be given additional training in recognising poultry expressions if required.

11.2 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 & 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 & 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 poultry. More specifically, the aim was to investigate the reliability of such a scale for laying hens and broilers kept under a variety of environmental conditions. Unfortunately the outbreak of Avian Influenza and the threat of its spread in the UK prevented completion of the work with broilers. The data collected were not sufficient for meaningful analysis, and this report therefore only presents the results obtained for laying hens.

11.3 METHODS

11.3.1 Development of Rating Scale

A first selection of qualitative descriptors of poultry behaviour was made on the basis of a previous FCP-study by Wang (2004). Using 10 experienced poultry workers as observers, she found three main dimensions of behavioural expression describing the variation between laying hens assessed in various tests under various environmental conditions. These dimensions are characterised by clusters of terms at both their positive and negative ends; from these clusters, a list of 23 terms was selected as a starting-point for this study. This list was sent to 6 research partners in the Welfare Quality poultry-subgroup for comments and amendments, of which, after repeated reminders, only 2 replied. Feedback from 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 28 descriptors.

The rating scale to be tested was construed by putting each of the 28 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, as the on-farm work was to be done in Germany, a German 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.

11.3.2 Reliability Testing

Assessments were carried out in 23 laying hen houses on 17 farms in different regions of Germany, with hens between 32 and 38 weeks of age. The farms differed with regard to farming practice (conventional/organic), housing system, group size, hen line and practice of beak trimming. A 'try out' assessment was made on one additional farm, but these data

were not included in the reliability analysis. The assessors were three females with different professional backgrounds (ethologist, agricultural student, animal caretaker), who were experienced in handling laying hens, and, to differing degrees, were knowledgeable about laying hen biology.

Before the first assessment, the three assessors 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. In the first three units observation times were between 3.5 and 14 minutes per observation point, but were shortened thereafter as specified in the instructions. The assessors wore either green or white overalls depending on the colour the animals were accustomed to. Assessors observed the same groups of birds 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). After finishing all observations, the assessors left the hen house 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 three assessors were measured with a ruler (i.e. the distance in millimetres between the minimum point 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 different terms to account for the variation between observed flocks of hens. The scores attributed to the 23 hen units on the two main components of the PCA were 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 scale, which are described below under 'Results'.

11.4 RESULTS

Table 9.1 shows the Kendall W values for each of the 28 qualitative descriptors on the rating scale. Setting an arbitrary minimum of 0.7 (Martin & Bateson, 1993), 20 terms out of 28 reached satisfactory inter-observer reliability (rounded up from 0.65).

TABLE 9.1 Kendall correlation coefficients for all descriptors (N = 3; df = 22).

Descriptor	Kendall's W	p-value	Descriptor	Kendall's W	p-value
Happy	.85	.000	Bored	.82	.000
Welfare overall	.85	.000	Distressed	.81	.000
Active	.84	.000	Scared	.77	.000
Positively occupied	.82	.000	Unsure	.72	.001
Comfortable	.77	.000	Tense	.70	.002
Confident	.76	.001	Frustrated	.70	.002
Friendly	.75	.001	Fearful	.68	.003
Content	.74	.001	Depressed	.68	.003
Relaxed	.73	.001	Unwell	.67	.003
Inquisitive	.63	.007	Nervous	.65	.005
Playful	.62	.008	Agitated	.65	.005
Calm	.57	.020	Vigilant/watchful	.51	.055
Coordinated as a group	.57	.021	Aggressive	.47	.097
Energetic	.51	.052	Compulsive	.44	.149

The assessors used the 28 terms in a very similar way. They distinguished between hen behavioural expressions of positive and negative mood (PC1), and between high and low levels of arousal in these moods (PC2). 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 Table 9.2.

To further adjust the rating scale for practical application, we reduced the number of descriptors from 28 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, and 2) had meanings without clear relevance to welfare. We also took out 'welfare overall' as this term includes non-animal environmental features. This reduction led to the removal of 6 out of 8 terms with a Kendall W value lower than 0.7 (inquisitive, playful, coordinated as a group, vigilant, aggressive, compulsive). Two terms with low W value (calm, energetic) stayed on the list because of their high loadings on PC2. Of the 16 terms characterising PC1 (see Table 9.2), 12 had Kendall W values above 0.7, and 3 between 0.6 and 0.7. The reduced QBA rating scale for poultry thus consisted of the following 20 terms: active, positively occupied, energetic, confident, comfortable, content, playful, relaxed, calm, happy, frustrated, depressed, fearful, scared, agitated, nervous, unsure, tense, bored, distressed.

We then checked the inter-observer reliability of this reduced rating scale by taking out the selected terms and their scores from the assessor files, conducting a PCA on these reduced files, and calculating Kendall W for its main components. Finally, we investigated the extent to which assessors' PC1 scores reflected their view of the hens' 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 Table 9.2.

TABLE 11.2 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=3, df=22, p<.001)	.85 / *.83*				.76 / *.73*		
% of variation explained	obs1 57 / *58*	obs2 65 / *70*	obs3 72 / *75*		obs1 17 / *19*	obs2 15 / *16*	obs3 10 / *11*
Main descriptors of PCA components	Content, comfortable, relaxed, confident, positively occupied, happy, calm Tense, distressed, nervous, unsure, frustrated, depressed, fearful, bored, scared				Energetic, agitated, active, playful, fearful, scared, nervous, positively occupied Calm, bored		
Correlation with 'welfare overall' (df=21; all p<0.001)	obs1 .93	obs2 .74	obs3 .96	mean .88			

Notes: roman values refer to the PCA of 33 terms, italic values to the PCA of 20 terms; the main descriptors of PCA components are valid for both analyses.

11.5 DISCUSSION

The aim of this study was to design a fixed rating scale of qualitative descriptors of hen behavioural expression and to test its inter-observer reliability. In designing this scale we preferred to start with a rather extensive list of 28 terms, and use the test results to remove less effective terms. Testing the rating scale with three assessors on 23 poultry farms showed that 20 out of 28 descriptors reached satisfactory reliability. However, more importantly, we found that Principal Component Analysis (PCA) of each assessor's scores resulted in remarkably similar clusters of descriptors at the positive and negative ends of the two principal components (PCs). From these clusters emerged a coherent pattern of poultry 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.

These findings suggest that poultry 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 analyses appear sufficiently robust to support this proposal; PC1 (based on the reduced 20-term PCA) explained a respectable 58–75% of the total variation between flocks (depending on the assessor), and showed good inter-observer reliability (0.83). Furthermore, it showed a very high mean correlation (0.88) to the assessors' score of 'welfare overall', which supports its direct relevance to poultry welfare.

Our preference for this approach was strengthened by the results of the 'Welfare Quality' QBA work with pigs and cattle. Here too, PCA revealed patterns of pig and cattle 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 reasonable to assume that our PCA results reflect a robust pattern of animal behavioural expression, which can reliably be used as an indicator of poultry welfare.

Further 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.

11.6 CONCLUSIONS

In conclusion, we suggest that the Qualitative Behaviour Assessment rating scale developed and tested in this study meets the required criteria of good inter-observer reliability and high poultry welfare relevance. This measure is highly feasible and easy to learn, and we recommend that it should be part of the WQ monitoring tool.

BIBLIOGRAPHY

Abeyesinghe, S.M., Nicol, C.J., Hartnell, S.J., Wathes, C.M. (2005) 'Can domestic fowl, Gallus gallus domesticus, show self-control?', *Animal Behaviour*, 70, pp. 1–11.

Abrahamsson, P., Tauson, R. and Elwinger, K. (1996b) 'Effects on production, health and egg quality of varying proportions of wheat and barley in diets of two hybrids of laying hens kept in different housing systems', *Acta Agriculturae Scandinavica, Section A: Animal Science*, 46, pp. 173–182.

Adams, A.W., Craig, J.V. and Bhagwat, A.L. (1978) 'Effects of flock size, age at housing, and mating experience on two strains of egg-type chickens in colony cages', *Poultry Science*, 57, pp. 48–53.

Akpobome, G.O. and Fanguy, R.C. (1992) 'Evaluation of cage floor systems for production of commercial broilers', *Poultry Science*, 71, pp. 274–280.

Allen, J. and Perry, G.C. (1975) 'Feather pecking and cannibalism in a caged layer flock', *British Poultry Science*, 16, pp. 441–445.

Altman, D.G. (1991) *Practical Statistics for Medical Research*, 1st ed. Chapman & Hall, London.

Anon (2006) *Defra Funded project AW0230 University of Bristol*, Report of the 2003 – 2005 survey of lameness in commercial broiler flocks.

Barnett, J.L. and Hemsworth, P.H. (1989) 'Fear of humans by laying hens in different tiers of a battery: behavioural and physiological responses', *British Poultry Science*, 30, pp. 497–502.

Barnett, J.L., Hemsworth, P.H., Hennessy, D.P., McCallum, T.H. and Newmann, E.A. (1994) 'The effects of modifying the amount of human contact on behavioural, physiological and production responses of laying hens', *Applied Animal Behaviour Science*, 41, pp. 87–100.

Barnett, J.L., Hemsworth, P.H. and Newman, E.A. (1992) 'Fear of humans and its relationship with productivity in laying hens at commercial farm', *British Poultry Science*, 33, pp. 699–710.

Bäumer, E. (1962) 'Lebensart des Haushuhns, dritter Teil: über seine Laute und allgemeine Ergänszungen, *Zeitschrift für Tierpsychologie*, 19, pp. 394–416.

Berg, C. and Sanotra, G.S. (2001) 'Kartlaggning av forekomsten av benfel hos svenska slaktkycklingar: en pilotstudie (with translation)', *Svensk Veterinartidning*, 53, p. 1.

Bestman, M.W.P. and Wagenaar, J.P. (2003) 'Farm level factors associated with feather pecking in organic laying hens', *Livestock Production Science*, 80, pp. 133–140.

Bilcik, B. and Keeling, L.J. (1999) 'Changes in feather condition in relation to feather pecking and aggressive behaviour in laying hens', *British Poultry Science*, 40, pp. 444–451.

Bilgili, S.F. (1999) 'Recent advances in electrical stunning', *Poultry Science*, 78, pp. 282–286.

Bjerrum, L., Pedersen, K. and Engberg, R.M. (2005) 'The influence of whole wheat feeding on salmonella infection and gut flora composition in broilers', *Avian Diseases*, 49, pp. 9–15.

Blair, R., Newberry, R.C. and Gardiner, E.E. (1993) 'Effects of lighting pattern and dietary tryptophan supplementation on growth and mortality in broilers', *Poultry Science*, 3, pp. 495–502.

Boissy, A. (1995) 'Fear and fearfulness in animals', *Quarterly Review of Biology*, 70(2), pp. 165–191

Boissy, A., Manteuffel, G., Jensen, M.B., Moe, R.O., Spruijt, B., Keeling, L.J., Winckler, C., Forkman, B., Dimitrov, I., Langbein, J., Bakken, M., Veissier, I. and Aubert, A. (2007) 'Assessment of positive emotions in animals to improve their welfare', *Physiology and Behaviour*, 92, pp. 375–397.

Bokkers, E.M. and Koene, P. (2004) 'Motivation and ability to walk for a food reward in fast and slow growing broilers to 12 weeks of age', *Behavioural Processes*, 67, pp. 121–130.

Bolhuis, J.E., Schouten, W.G.P., Schrama, J.W. and Wiegant, V.M. (2005) 'Behavioural development of pigs with different coping characteristics in barren and substrate-enriched housing conditions', *Applied Animal Behaviour Science*, 93(3–4), pp. 213–228.

Bradshaw, R.H., Kirkden, R.D. and Broom, D.M. (2002) 'A review of the aetiology and pathology of leg weakness in broilers in relation to welfare', *Avian and Poultry Biology Reviews*, 13(2), pp. 45–103.

Bray, T.S. and Lynn, N.J. (1986) 'Effects of nutrition and drinker design on litter condition and broiler performance', *British Poultry Science*, 151.

Bright, A., Jones, T.A. and Dawkins, M.S. (2006) 'A non-intrusive method of assessing plumage condition in commercial flocks of laying hens', *Animal Welfare*, 15, pp. 113–118.

Bruder, C. (1988) 'Défauts de présentation des carcasses de poulet', *Doc Éleveur*, juillet-août, pp. 34–36.

Butterworth, A. (2000) 'Infectious components of broiler lameness: a review', *World's Poultry Science Journal*, 327–352.

Butterworth, A. (2000) 'Welfare of broilers in organic systems', *AWSELVA*.

Butterworth, A. (2001) *Infectious Components of Broiler Lameness*. PhD Thesis, University of Bristol.

Butterworth, A., Reeves, N.A., Harbour, D., Werret, G. and Kestin, S.C. (2001) 'Molecular typing of strains of Staphylococcus aureus isolated from bone and joint lesions in lame broilers by Random Amplification of Polymorphic DNA', *Poultry Science*, 80, pp. 1339–1343.

Butterworth, A., Weeks, C.A., Crea, P.R. and Kestin, S.C. (2002) 'Dehydration and lameness in a broiler flock', *Animal Welfare*, 11, pp. 89–94.

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.

Chamblee, T.N., Morgan, G.W. and Schultz, C.D. (1989) 'Effect of refeeding following short-term deprivation of feed or water, or both, on selected physiological parameters for broiler chickens', *Poultry Science*, 18, pp. 1619–1623.

Chapuis, H., Boulay, M., Retailleau, J.-P., Arnould, C., Mignon-Grasteau, S., Berri, C., Besnard, J., Coudurier, B. and Faure, J.-M. (2003) 'Sélection d'une souche de poulet label contre le picage: bilan après trois générations de sélection au picomètre', *5e Journées de la Recherche Avicole*, 26-27 March, Tours, France, pp. 363–366.

Clark, S., Hansen, G., Mc Lean, P., Bond, P., Wakeman, W., Meadows, R. and Buda, S. (2002) 'Pododermatitis in turkeys', *Avian Diseases*, 46, pp. 1038–1044.

Classen, H.L. and Riddell, C. (1989) 'Photoperiodic effects on performance and leg abnormalities in broiler chickens', *Poultry Science*, pp. 873–879.

Classen, H.L. and Riddell, C. (1990) 'Early growth rate and lighting effects on broiler skeletal disease', *Poultry Science*, 69(suppl.), p. 35 (abstr.).

Classen, H.L., Riddell, C. and Robinson, F.E. (1991) 'Effects of increasing photoperiod length on performance and health of broiler chickens', *British Poultry Science*, pp. 21–29.

Colas, M. (2005) 'Le bien-être des poulets de chair: Quels critères pertinents pour son évaluation en routine? Mémoire de fin d'études', ESITPA. (Now published as: Allain, V., Mirabito, L., Arnould, C., Colas, M., Le Bouquin, S., Lupo, C. and Michel, V. (2009) 'Skin lesions in broiler chickens measured at the slaughterhouse: relationships between lesions and between their prevalence and rearing factors', *British Poultry Science*, 50, pp. 407–417.)

Collias, N.E. (1987) 'The vocal repertoire of the Red Junglefowl: a spectrographic classification and the code of communication', *Condor*, 89, pp. 510–524.

Corr, S.A., Maxwell, M.J., Gentle, D. and Bennett, D. (2003) 'Preliminary study of joint disease in poultry by the analysis of synovial fluid', *Veterinary Record*, 152, pp. 549–554.

Council of the European Union (1993) 'Council Directive 93/119/EC of 22 December 1993 on the protection of animals at the time of slaughter or killing', *Official Journal of the European Communities*, L 340, pp. 21–34.

Dalin, R. (2003) *Statistik för vetenskapligt arbete*. Mittuniversitetet, Östersund.

Danbury, T.C., Weeks, C.A., Chambers, J.P., Waterman-Pearson, A.E. and Kestin, S.C. (2000) 'Self-selection of the analgesic drug carprophen by lame broiler chickens', *Veterinary Record*, 146, pp. 307–311.

Dawkins, M.S. (1989) 'Time budgets of Red Junglefowl as a baseline for the assessment of welfare in domestic fowl', *Applied Animal Behaviour Science*, 24, pp. 77–80.

Dawkins, M.S., Donnelly, C. and Jones, T.A. (2004) 'Chicken welfare is influenced more by housing conditions than by stocking density', *Nature*, 427, pp. 342–344.

DECCA (2001) *Statement on Laying Hens*. Danish Ethical Council Concerning Animals, Copenhagen.

Deschutter, A. and Leeson, S. (1986) 'Feather growth and development', *World's Poultry Science Journal*, 42, pp. 259–267.

Devriese, L.A., Cauwerts, K., Hermans, K., Wood, A.M. (2002) 'Enterococcus cecorum septicemia as a cause of bone and joint lesions resulting in lameness in broiler chickens', *Vlaams Diergeneeskundig Tijdschrift*, 71, pp. 219–221.

Dobrzanski, Z. and Bialas, W. (1993) 'Effectiveness of the floor heating system in broiler houses', in: *Proceedings of International Livestock Environment Symposium IV*, ASAE, Warwick.

Donaldson, T.M., Newberry, R.C., Spinka, M. and Cloutier, S. (2002) 'Effects of early play experience on play behaviour of piglets after weaning', *Applied Animal Behaviour Science*, 79(3), pp. 221–231.

Duff, S.R.I. and Thorp, B.H. (1985) 'Abnormalangulation/torsion of the pelvic appendicular skeleton in broiler fowl: morphological and radiolgical findings', *Research in Veterinary Science*, pp. 313–319.

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.

Edwards, H.M. and Sorensen, P. (1987) 'Effects of short fasts on the development of tibial dyschondroplasia in chickens', *Journal of Nutrition*, pp. 194–200.

Ekstrand, C., Algers, B. and Svedberg, J. (1997) 'Rearing conditions and foot-pad dermatitis in Swedish broiler chickens', *Preventive Veterinary Medicine*, pp. 167–174.

Ekstrand, C., Carpenter, T.E., Anderson, I. and Algers, B. (1998) 'Prevalence and control of foot-pad dermatitis in broilers in Sweden', *British Poultry Science*, 39, pp. 318–324.

El-Lethey, H., Jungi, T.W. and Huber-Eicher, B. (2001) 'Effects of feeding corticosterone and of housing conditions on feather pecking in laying hens (Gallus gallus domesticus)', *Physiology and Behavior*, 73, pp. 243–251.

Elfadil, A.A., Vaillancourt, J.P., Meek, A.H. (1996) 'Impact of stocking density, breed, and feathering on the prevalence of abdominal skin scratches in broiler chickens', *Avian Diseases*, 40, pp. 546–552.

Elwinger, K. (1995) 'Broiler production under varying population densities: a field study', *Archiv für Geflügelkunde*, 59, pp. 209–215.

Engel, B. and Reenen, K. van (2006) Reliability of Measures of Farm Animal Welfare: Intra- and Inter-observer Variability, note for EC Welfare Project, WP 2.1, 3.13.2006.

Estep, D.Q. and Hetts, S. (1992) 'Interactions, relationships, and bonds: the conceptual basis for scientist-animal relations', in: H. Davis and A.D. Balfour (eds) *The Inevitable Bond: Examining Scientist–Animal Interactions*, CAB International, Cambridge, pp. 6–26.

European Commission (1991) 'Commission Regulation (EEC) No 1538/91 of June 5, 1991, introducing detailed rules for implementing Regulation (EEC) No. 1906/90 on certain marketing standards for poultry', *Official Journal of the European Union*, L 143, pp. 11–22.

European Commission (2000) *The Welfare of Chickens Kept for Meat Production (Broilers)*, Report of the Scientific Committee on Animal Health and Animal Welfare, published online <http://www.cordis.lu>.

FAWC (1992) *Report on the Welfare of Broiler Chicken*. Farm Animal Welfare Council, Surbiton.

FAWC (1998) *Report on the Welfare of Broiler Breeders*. Farm Animal Welfare Council, Surbiton.

Fleiss, J.L. (1971) 'Measuring nominal scale agreement among many raters', *Psychological Bulletin*, 76, pp. 378–382.

Fölsch, D.W. (1981) 'Das Verhalten von Legehennen in unterschiedlichen Haltungssystemen unter Berücksichtigung der Aufzuchtmethoden', in: D.W. Fölsch

and K. Vestergaard (eds) *Das Verhalten von Hühnern*, Birkhäuser Verlag, Basel, pp. 9–114.

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. 531–565.

Fraisse, F, Cockrem, J F (2006) 'Corticosterone and fear behaviour in white and brown caged laying hens', *British Poultry Science*, 47(2): 110-119

Frankenhuis, M.T., Vertommen, M.H. and Hemminga, H. (1991) 'Influence of claw clipping, stocking density and feeding space on the incidence of scabby hips in broilers', *British Poultry Science*, 32, pp. 227–230.

Fraser, A.F. and Broom, D.M. (1990) *Farm Animal Behaviour and Welfare*. Bailiére Tindall, London.

Freire, R., Walker, A. and Nicol, C.J. (1999) 'The relationship between trough height, feather cover and behaviour of laying hens in modified cages', *Applied Animal Behaviour Science*, 63, pp. 55–64.

Freire, R., Wilkins, L.J., Short, F. and Nicol, C.J. (2003) 'Behaviour and welfare of individual laying hens in non-cage systems', *British Poultry Science*, 44, pp. 22–29.

Frijtag, J.C. von, Reijmers, L.G.J.E., Van der Harst, J.E., Leus, I.E., Van den Bos, R. and Spruijt, B.M. (2000) 'Defeat followed by individual housing results in long-term impaired reward- and cognition-related behaviours in rats', *Behavioural Brain Research*, 117(1–2), p. 1.

Gentle, M.J. (2001) 'Attentional shifts alter pain perception in chickens', *Animal Welfare*, 10, pp. 187–236.

Gentle, M.J. and Hunter, L.N. (1990) 'Physiological and behavioural responses associated with feather removal in Gallus gallus var domesticus, *Research in Veterinary Science*, 50, pp. 95–101.

Gerritzen, M.A., Lambooij, E., Hillebrand, S.J.W., Lankhaar., J.A.C. and Pieterse, C. (2000) 'Behavioural responses of broilers to different gaseous atmospheres', *Poultry Science*, 79, pp. 928–933.

Goldstein, H. (1995) *Multilevel Statistical Models*, Kendalls Library of Statistics. Arnold, London.

Gomis, S.M., Watts, T., Riddell, C., Potter, A.A. and Allan, B.J. (1997) 'Experimental reproduction of Esherichis coli cellulitis and septicemia in broiler chickens', *Avian Diseases*, pp. 234–240.

Gordon, R.F., Jordan, F.T.W. and Pattison, M. (1996) *Poultry Diseases*, 4th rev. edn, Saunders (W.B.), London.

Gosling, S.D. (2001) 'From mice to men: what can we learn about personality from animal research?', *Psychological Bulletin*, 127, pp. 45–86.

Graml, C., Waiblinger, S., Niebuhr, K. (2008a) 'Validation of tests for on-farm assessment of the hen–human relationship in non-cage systems', *Applied Animal Behaviour Science*, 111, pp. 301–310.

Graml, C., Niebuhr, K., Waiblinger, S. (2008b) 'Reaction of laying hens to humans in the home or a novel environment', *Applied Animal Behaviour Science*, 113, pp. 98–109.

Grashorn, M. and Kutritz, B. (1991) 'Effect of stocking density on performance of modern broiler breeds', *Archiv für Geflügelkunde*, 2, pp. 84–90.

Greene, J.A. (1985) 'A contact dermatitis of broilers: clinical and pathological findings', *Avian Pathology*, 14, pp. 31–63.

Greene, J.A., McCracken, R. and Evans, R.T. (1985) 'A contact dermatitis of broilers: clinical and pathological findings', *Avian Pathology*, 14, pp. 23–28.

Gregory, N.G. (1998) *Animal Welfare and Meat Science*. CABI Publishing, Wallingford.

Gregory, N.G. and Austin, S.D. (1992) 'Causes of trauma in broilers arriving dead at poultry processing plants', *Veterinary Record*, 131(22), pp. 501–503.

Guesdon, V. (2004) *Etude comparative de poules pondeuses épointées ou non élevées en cages standard ou aménagées: estimation multicritères du bien-être*. Thèse de Docteur de l'Université de Rennes I.

Gunnarsson, S., Odén, K., Algers, B., Svedberg, J. and Keeling, L. (1995) *Poultry Health and Behaviour in a Tiered System for Loose Housed Layers*, Rapport 35. Swedish University of Agricultural Sciences, Skara.

Gunnarsson, S., Keeling, L.J. and Svedberg, J. (1999) 'Effect of rearing factors on the prevalence of floor eggs, cloacal cannibalism and feather pecking in commercial flocks of loose housed laying hens', *British Poultry Science*, 40, pp. 12–18.

Gunnarsson, S., Algers, B. and Svedberg, J. (2000) 'Description and evaluation of a scoring system of clinical health in laying hens', in: S. Gunnarsson *Laying Hens in Loose Housing Systems*, Doctoral thesis, Swedish University of Agricultural Sciences, Uppsala, pp. I1-I18.

Gyger, M., Marler, P. and Pickert, R. (1987) 'Semantics of an avian alarm call system: the male domestic fowl, Gallus gallus domesticus', *Behaviour*, 102, pp. 13–40.

Hafez, H.M., Mazaheri, A., Prusas, C., Böhland, K., Pöppel, M. and Schulze, D. (2001) 'Aktuelle Geflügelkrankheiten bei Legehennen im Zusammenhang mit alternativen Haltungssystemen', *Tieärztliche Praxis*, 29(G), pp. 168–174.

Hall, A.L. (2001) 'The effect of stocking density on the welfare and behaviour of broiler chickens reared commercially', *Animal Welfare*, 10, pp. 23–40.

Häne, M., Huber-Eicher, B. and Fröhlich, E. (2000) 'Survey of laying hen husbandry in Switzerland', *World's Poultry Science Journal*, 56, pp. 21–31.

Harms, R.H. and Simpson, C.F. (1975) 'Biotin deficiency as a possible cause of swelling and ulceration of footpads', *Poultry Science*, pp. 1711–1713.

Harms, R.H., Damron, B.L. and Simpson, C.F. (1977) 'Effect of wet litter and supplemental biotin and/or whey on the production of foot pad dermatitis in broilers', *Poultry Science*, pp. 291–296.

Haye, U. and Simons, P.C.M. (1978) 'Twisted legs in broilers', *British Poultry Science*, pp. 549–557.

Hegelund, L., Sørensen, J.T. and Johansen, N.F. (2003) Developing a welfare assessment system for use in commercial organic egg production, *Animal Welfare*, 12, pp. 649–653.

Hemminga, H. and Vertommem, M.H. (1985) 'Faecal burn-spots upon the breast-skin of broiler chickens', *Pluimveehouderij*, pp. 12–14.

Herremans, M. and Decuypere, E. (1987) 'A new approach to recording plumage deterioration', *British Poultry Science*, 28, pp. 461–470.

Huber-Eicher, B. and Sebö, F. (2001) 'The prevalence of feather pecking and development in commercial flocks of laying hens', *Applied Animal Behaviour Science*, 74, pp. 223–231.

Huchzermeyer, F.W. and De Ruyck, A.M.C. (1986) 'Causes and prevention of broiler ascites', *(SAPA) Poultry Bulletin*, August, pp. 346–340.

Hughes, B.O. and Duncan, I.J.H. (1972) 'The influence of strain and environmental factors upon feather pecking and cannibalism in fowls', *British Poultry Science*, 13, pp. 525–547.

Hurwitz, S. (1992) 'The role of vitamin D in poultry bone biology', in: C.C. Whitehead (ed.) *Bone Biology and Skeletal Disorders in Poultry*, Oxford: Carfax, pp. 87–102.

Jones, R.B. (1985a) 'Fear responses of individually-caged laying hens as a function of cage level and aisle', *Applied Animal Behaviour Science*, 14(1), pp. 63–74

Jones, R.B. (1985b) 'Fearfulness of hens caged individually or in groups in different tiers of a battery and the effects of translocation between tiers', *British Poultry Science*, 26, pp. 399–408.

Jones, R.B. (1987) 'Fearfulness of caged laying hens: the effects of cage level and type of roofing', *Applied Animal Behaviour Science*, 17, pp. 171–175.

Jones, R.B. (1988) 'Repeatability of fear ranks among adult laying hens', *Applied Animal Behaviour Science*, 19(3–4), pp. 297–304.

Jones, R.B. (1993) 'Reduction of the domestic chick's fear of human beings by regular handling and related treatments', *Animal Behaviour*, 46, pp. 991–998.

Jones, R.B. (1994) 'Regular handling and domestic chick's fear of human beings: generalisation of response', *Applied Animal Behaviour Science*, 42, pp. 129–143.

Jones, R.B. (1995) 'Ontogeny of response to humans in handled and non-handled female domestic chicks', *Applied Animal Behaviour Science*, 42, pp. 261–269.

Jones, R.B. (1996) 'Fear and adaptability in poultry: insights, implications and imperatives', *World's Poultry Science Journal*, 52, pp. 131–174.

Jordan, F.T.W., Pattison, M., Alexander, D.J. and Faragher, T. (eds) (2001) *Poultry Diseases*, 5th edn. WB Saunders, London.

Julian, R.J. (2005) 'Production and growth related diorders and other metabolic diseases of poultry: a review', *Veterinary Journal*, 169, pp. 350–369.

Keer-Keer, S., Hughes, B.O., Hocking, P.M., Jones, R.B. (1996) 'Behavioural comparison of layer and broiler fowl: measuring fear responses', *Applied Animal Behaviour Science*, 49(4), pp. 321–333.

Keppler, C., Schubbert, A. and Knierim, U. (2005) 'Plumage condition and injuries as indicators of welfare in layers: which scoring-method is suitable?', *Animal Science Papers and Reports*, 23(Suppl. 1), pp. 273–274.

Keppler, C., Lange, K., Strobel, E. and Fölsch, D.W. (2001) 'A comparative study of the influence of breed on feather pecking and cannibalism in laying hens in alternative rearing and husbandry systems including feeding aspects', in: H. Oester and C. Wyss (eds) *Proceedings of the 6th European Symposium on Poultry Welfare, Zollikofen, 1–4 September*, pp. 289–291.

Kestin, S.C., Knowles, T.G., Tinch, A.E. and Gregory, N.G. (1992) 'Prevalence of leg weakness in broiler chickens and its relationship with genotype', *Veterinary Record*, 131 pp. 190–194.

Kestin, S.C., Adams, J.M. and Gregory, N.G. (1994) 'Leg weakness in broilers, a review of studies using gait scoring', in: *European Poultry Conference*, pp. 203–206.

Kestin, S.C., Su, G. and Sorensen, P. (1999) 'Different commercial broiler crosses have different susceptibilities to leg weakness', *Poultry Science*, pp. 1085–1090.

Kestin, S.C., Gorden, S., Su, G. and Sorensen, P. (2001) 'Relationships in broiler chickens between lameness, live weight, growth rate and age', *Veterinary Record*, 148 pp. 195–197.

Kjaer, J.B. and Sørensen, P. (2002) 'Feather pecking and cannibalism in free-range laying hens as affected by genotype, dietary level of methionine plus cystine, light intensity during rearing and age at first access to the range area', *Applied Animal Behaviour Science*, 76, pp. 21–39.

Knowles, T.G., Ball, R.C., Warriss, P.D. and Edwards, J.E. (1996) 'A survey to investigate potential dehydration in slaughtered broiler chickens', *British Veterinary Journal*, 152, pp. 307–314.

Knowles, T.G., Warris, P.D., Brown, S.N., Edwards J.E. and Mitchell, M.A. (1995) 'Response of broilers to deprivation of food and water for 24 hours', *British Veterinary Journal*, 151, pp. 197–202.

Koike, T.I., Pryor, L.R. and Neldon, H.L. (1983) 'Plasma volume and electrolytes during progressive water deprivation in chickens (Gallus domesticus)', *Comparative Biochemistry and Physiology*, 74A, pp. 83–87.

Kristensen, H.H, Perry, G.C., Prescott, N.B., Ladewig, J., Ersboll, A.K. and Wathes, C. M. (2006) 'Leg health and performance of broiler chickens reared in different light environments', *British Poultry Science*, 47, pp. 257–263.

Kruijt, J.P. (1964) 'Ontogeny of social behaviour in Burmese red junglefowl (Callus gallus spadiceus) Bonnaterre', *Behaviour*, 12(Suppl 2), pp. 1–201.

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.

Le Bihan-Duval, E., Beaumont, C. and Colleau, J.J. (1996) 'Genetic parameters of the twisted legs syndrome in broiler chickens', *Genetics Selection Evolution*, 28, pp. 177–195.

Leterrier, C. and Nys, Y. (1992) 'Clinical and anatomical differences in varus and valgus deformities of chick limbs suggest different aetio-pathogenesis', *Avian Pathology*, pp. 429–442

Lilburn, M.S., Lauterio, T.J., Ngiam-Rilling, K. and Smith, J.H. (1989) 'Relationships among mineral balance in the diet, early growth manipulation, and incidence of tibial dyschondroplasia in different strains of meat type chickens', *Poultry Science*, pp. 1263–1273.

Lindblad, J. (2006) Veterinarian, Svensk Fågel, Personal message, 3 February.

Martin, P. and Bateson, P. (1993) *Measuring Behaviour*, 2nd edn. Cambridge University Press, Cambridge.

Martland, M.F. (1985) 'Ulcerative dermatitis in broiler chickens: the effects of wet litter', *Avian Pathology*, pp. 353–364.

Martrenchar, A., Morisse, J. P., Huonnic, D. and Cotte, J. P. (1997) 'Influence of stocking density on some behavioural, physiological and productivity traits of broilers', *Veterinary Research*, 28, pp. 473–480.

Martrenchar, A., Boilletot, E, Huonnic, D. and Pol, F. (2002) 'Risk factors for foot-pad dermatitis in chicken and turkey broilers in France, *Preventive Veterinary Medicine*, 52, pp. 213–226.

May, K.N. and Hamdy, M.K. (1966) 'Bruising of poultry: a review', *World's Poultry Science Journal*, 22, pp. 316–322.

May, J.D., Lott, B.D. and Deaton, J.W. (1990) 'The effect of light and environmental temperature on broiler digestive tract contents after feed withdrawal', *Poultry Science*, 69, pp. 1681–1684.

Mayne, R.K. (2005) 'A review of the aetiology and possible causative factors of foot pad dermatitis in growing turkeys and broilers', *World's Poultry Science Journal*, 61, pp. 256–267.

McIlroy, S.G., Goodall, E.A. and McMurray, C.H. (1987) 'A contact dermatitis of broilers: epidemiological findings', *Avian Pathology*, pp. 93–105.

McLean, J.A., Savory, C.J. and Sparks, N.H.C. (2002) 'Welfare of male and female broiler chickens in relation to stocking density, as indicated by performance, health and behaviour', *Animal Welfare*, 11, pp. 55–73.

McNamee, P.T. and Smyth, J.A. (2000) 'Bacterial chondronecrosis with osteomyelitis (femoral head necrosis) of broiler chickens: a review', *Avian Pathology*, 29, pp. 253–270.

McNamee, P.T., McCullagh, J.J., Thorp, B.H., Ball, H.J., Graham, D., McCullough, S.J., McConaghy, D. and Smyth, J.A. (1998) 'Study of leg weakness in two commercial broiler flocks', *Veterinary Record*, pp. 131–135.

McNamee, P.T., McCullagh, J.J., Rodgers, J.D., Thorp, B.H., Ball, H.J., Connor, T.J., McConaghy, D. and Smyth, J.A. (1999) 'Development of an experimental model of bacterial chondronecrosis with osteomyelitis in broilers following exposure to Staphylococcus aureus by aerosol, and inoculation with chicken anaemia and infectious bursal disease viruses', *Avian Pathology*, 28, pp. 26–35.

Meldrum, R.J., Griffiths, J.K., Smith, R.M.M. and Evans, M.R. (2005) 'The seasonality of human campylobacter infection and Campylobacter isolates from fresh, retail chicken in Wales', *Epidemiology and Infection*, 133, pp. 49–52.

Mench, J.A., Garner, J.P. and Falcone, C. (2001) 'Behavioural activity and its effects on leg problems in broiler chickens', in: H. Oester and C. Wyss (eds) *Proceedings of the 6th European Symposium on Poultry Welfare, Zollikofen, Switzerland*, pp. 152–156.

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.

Mirabito, L., Arnould, C., Colas, M. and Michel, V. (2006) 'Relationship between different types of skin lesions in broilers reared commercially. XII European Poultry Conference, 10–14 September', *World's Poultry Science Journal*, 62, S., pp. 582–583.

Nairn, M.E. and Watson, A.R.A. (1972) 'Leg weakness in broilers: a clinical and pathological characterisation', *Australian Veterinary Journal*, pp. 645–656.

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., Hunt, J.R. and Gardiner, E.E. (1988) 'Influence of light intensity on behavior and performance of broiler chickens', *Poultry Science*, pp. 1020–1025.

Newberry, R.C., Wood-Gush, D.G.M., Hall, J.W. (1988) 'Playful behavior of piglets', *Behavioural Processes*, 17(3), pp. 205–216.

Nicol, C.J., Pötzsch, C., Lewis, K. and Green, L.E. (2003) 'Matched concurrent case-control study of risk factors for feather pecking in hens on free-range commercial farms in the UK', *British Poultry Science*, 44, pp. 515–523.

Niebuhr, K., Zaludik, K., Raubek, J., Gruber, B. and Waiblinger, S. (2007) Farmers' Attitudes, Daily Working and Handling Practices on Grower and Laying Hen Farms and their Relationship with Hen Welfare, Report for the EEC-project Welfare Quality®.

Nijdam, E., Arens, P., Lambooij, E., Decuypere, E. and Stegeman, J.A. (2004) 'Factors influencing bruises and mortality of broilers during catching, transport and lairage', *Poultry Science*, 83, pp. 1610–1615.

Nijdam, E., Zailan, A.R.M., van Eck, J.H.H., Decuypere, E. and Stegeman, J.A. (2006) 'Pathological features in dead on arrival broilers with special reference to heart disorders', *Poultry Science*, 85, pp. 1303–1308.

Oester, H. (1994) 'Sitzstangenformen und ihr Einfluß auf die Entstehung von Fußballengeschwüren bei Legehennen', *Archiv für Geflügelkunde*, 58, pp. 239–244.

Padron, M. (1990) 'Salmonella typhimurium outbreak in broiler chicken flocks in Mexico', *Avian Diseases*, pp. 221–223.

Pagazaurtundua, A. and Warriss, P.D. (2006) 'Measurements of footpad dermatitis in broiler chickens at processing plants', *Veterinary Record*, 158, pp. 679–682.

Pattison, M. (1987) 'Problems of diarrhoea and wet litter in meat poultry' in: W. Haresign and D.J.A. Cole (eds) *Recent Advances in Animal Nutrition*, Butterworths, London.

Pattison, M. (1992) 'Impacts of bone problems on the poultry meat industry', in: C.C. Whitehead (ed.) *Bone Biology and Skeletal Disorders in Poultry*, Oxford: Carfax.

Petitte, J.N., Hawes, R.O. and Gerry, R.W. (1983) 'The influence of cage versus pen management of broiler breeder hens on subsequent performance of cage reared broilers', *Poultry Science*, 62, pp. 1241–1246.

Poulus, P.W., Reiland, S., Elwinger, K. and Olsson, S.E. (1978) 'Skeletal lesions in the broiler, with special reference to dyschondroplasia (osteochondrosis). Pathology, frequency and clinical significance in two strains of bird on high and low energy feed', *Acta Radiologica*, 358, pp. 229–275.

Raj, M. (1998) 'Welfare during stunning and slaughter of poultry', *Poultry Science*, 77, pp. 1815–1819.

Raubek, J., Niebuhr, K. and Waiblinger, S. (2007) 'Development of on-farm methods to assess the animal-human relationship in laying hens kept in non-cage systems', *Animal Welfare*, 16, pp. 173–175.

Rauw, W.M. (1998) 'Undesirable side-effects to selection for high production efficiency', *Livestock Production Science*, 56, pp. 15–33.

Reece, F.N. and Deaton, J.W. (1971) 'Some observations on conditions in windowless poultry houses after loss of ventilation', *Poultry Science*, pp. 1885–1887.

Renden, J.A., Bilgili, S.F., Lien, R.J. and Kinaid, S.A. (1991) 'Live performance and yields of broilers provided various lighting schedules', *Poultry Science*, pp. 2055–2062.

Riddell, C. (1975) 'Studies on the pathogenesis of tibial dyschondroplasia in chickens: II. Growth rate of long bones', *Avian Diseases*, pp. 490–496.

Riddell, C. (1976) 'Selection of broiler chickens for high and low incidence of tibial dyschondroplasia with observations on spodylolysthesis and twisted legs (perosis)', *Poultry Science*, pp. 145–151.

Riddell, C. (1983) 'Pathology of the skeleton and tendons of broiler chickens reared to roaster weights: I. Crippled chickens', *Avian Diseases*, pp. 950–962.

Riddell, C. (1992) 'Non-infectious skeletal disorders of poultry: an overview', in: C.C. Whitehead (ed.) *Bone Biology and Skeletal Disorders in Poultry*, Oxford: Carfax, pp 119–145.

Riddell, C. and Springer, R. (1985) 'An epizootiological study of acute sudden death syndrome and leg weakness in broiler chickens in western Canada', *Avian Diseases*, pp. 90–102.

Robinson, F.E., Classen, H.L., Hanson, J.A. and Onderka, D.K. (1992) 'Growth performance, feed efficiency and the incidence of skeletal and metabolic disease in full-fed and feed restricted broiler and roaster chickens', *Journal of Applied Poultry Research*, pp. 33–41.

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.

Sainsbury, D.W.B. (1993) 'Poultry environment housing and hygiene', in: M. Pattison (ed.) *The Health of Poultry*, Longman Scientific and Technical, Harlow, pp. 23–69.

Sandilands, V., McGovern, R., Savory, C.J. (2003) 'Objectively measuring broiler walking style using a force plate', in: V. Ferrante (ed.) *Proceedings of the 37th International Congress of the ISAE*, Fondazione Iniziative Zooprofilattiche e Zootecniche, Brescia, p. 61.

Sanotra, G.S. (2000) *Leg Problems in Broilers: A Survey of Conventional Production Systems in Denmark*. Dyrenes Beskyttelse, Frederiksberg.

Sanotra, G.S., Berg, C. and Lund, J.D. (2003) 'A comparison between leg problems in Danish and Swedish broiler production', *Animal Welfare*, 12, pp. 677–683.

Savory, C.J. (1995) 'Feather pecking and cannibalism', *World's Poultry Science Journal*, 51, pp. 215–219.

Schenk, P.M., Meysser, F.M. and Limpens, H.J.G. (1983) 'Gakeln als Indikator für Frustration bei Legehennen', *KTBL- Schrift*, 299, pp. 65–80.

Scholz, B., Rönchen, S., Hamann, H., Hewicker-Trautwein, M., Distl, O. (2008) 'Keel bone condition in laying hens: a histological evaluation of macroscopically assessed keel bones', *Berliner und Münchener Tierärztliche Wochenschrift*, 121, pp. 89–94.

Sefton, A.F. (1976) 'The interactions of cage size, cage level, social density, fearfulness, and production of single comb white leghorns', *Poultry Science*, pp. 1922–1926.

Simonsen, H.B., Vestergaard, K. and Willeberg, P. (1980) 'Effect of floor type and density on integument of egg-layers', *Poultry Science*, 59, pp. 2202–2206.

Sorensen, P. and Kestin, S.C. (1999) 'The effect of photoperiod:scotoperiod on leg weakness in broiler chickens', *Poultry Science*, pp. 336–342.

Sorensen, P., Su, G. and Kestin, S.C. (2000) 'Effects of age and stocking density on leg weakness in broiler chickens', *Poultry Science*, 79, pp. 864–870.

Sparrey, J.M., and Kettlewell, P.J. (1994) 'Shackling of poultry: is it a welfare problem?', *World's Poultry Science Journal*, 50, pp. 167–176.

Spinka, M., Newberry, R.C. and Bekoff, M. (2001) 'Mammalian play: training for the unexpected', *Quarterly Review of Biology*, 76(2), pp. 141–168.

Spruijt, B.M., van den Bos, R. and Pijlman, F.T.A. (2001) 'A concept of welfare based on reward evaluating mechanisms in the brain: anticipatory behaviour as an indicator for the state of reward systems', *Applied Animal Behaviour Science*, 72(2), pp. 145–171.

Stephenson E.L., Bezanson J.M. and Hall C.F. (1960) 'Factors affecting the incidence and severity of a breast blister condition in broilers', *Poultry Science*, 39, pp. 1520–1524.

Sturkie, P.D. (ed.) (1986) *Avian Physiology*, 4th edn. New York: Springer-Verlag.

Su, G., Sorensen, P. and Kestin, S.C. (1999) 'Meal feeding is more effective than early feed restriction at reducing the prevalence of leg weakness in broiler chickens', *Poultry Science*, pp. 949–955.

Su, G., Sorensen, P. and Kestin, S.C. (2000) 'A note on the effects of perches and litter substrate on leg weakness in broiler chickens', *Poultry Science*, 79, pp. 1259–1263.

Tauson, R. and Abrahamsson, P. (1996) 'Foot and keel bone disorders in laying hens. Effects of artificial perch material and hybrid', *Acta Agriculturae Scandinavica*, Section A: Animal Science, 46, pp. 239–246.

Tauson, R. and Svensson, S.A. (1980) 'Influence of plumage condition on the hen`s feed requirement', *Swedish Journal of Agricultural Research*, 10, pp. 35–39.

Tauson, R., Ambrosen, T. and Elwinger, K. (1984) 'Evaluation of procedures for scoring the integument of laying hens: independent scoring of plumage condition', *Acta Agriculturae Scandinavica, Section A: Animal Science*, 34, pp. 400–408.

Tauson, R., Kjaer, J., Maia, G., Cepero, R. and Holm, K.-E. (2005) 'Applied scoring of integument and health in laying hens', *Animal Science Papers and Reports*, 23 (Suppl. 1), pp. 153–159.

Thorp, B.H. (1992) 'Abnormalities in the growth of leg bones', in: C.C. Whitehead (ed.) *Bone Biology and Skeletal Disorders in Poultry*, Oxford: Carfax, pp. 147–166.

Thorp, B.H. (1996) 'Skeletal disorders in fowl: a review', in: F.T.W. Jordan, M. Pattison, D.J. Alexander and T. Faragher (eds) (2001) *Poultry Diseases*, 5th edn. WB Saunders, London, pp. 290–305.

Tucker, S.A. and Walker, A.W. (1992) 'Hock burn in broilers', in: P.C. Garnsworthy, W. Haresign and D.J.A. Cole (eds) *Recent Advances in Animal Nutrition*, Butterworth Heinemann, Oxford, pp. 33–50.

Van den Bos, R., Meijer, M.K., van Renselaar, J.P., van der Harst, J.E. and Spruijt, B.M. (2003) 'Anticipation is differently expressed in rats (Rattus norvegicus) and domestic cats (Felis silvestris catus) in the same Pavlovian conditioning paradigm', *Behavioural Brain Research*, 141(1), pp. 83–89.

Van der Harst, J.E., Baars, A.M. and Spruijt, B.M. (2005) 'Announced rewards counteract the impairment of anticipatory behaviour in socially stressed rats', *Behavioural Brain Research*, 161(2), pp. 183–189.

Vestergaard, K., Kruijt, J.P. and Hogan, J.A. (1993) 'Feather pecking and chronic fear in groups of red junglefowl: their relations to dustbathing, rearing environment and social status', *Animal Behaviour*, 45, pp. 1127–1140.

Vinke, C.M., Van den Bosch, R. and Spruijt, B.M. (2004) 'Anticipatory activity and stereotypical behaviour in American mink (Mustela vison) in three housing systems differing in the amount of enrichments', *Applied Animal Behaviour Science*, 89(1–2), pp. 145–161.

Vits, A., Weitzenburger, D. and Distl, O. (2005) 'Vergleich verschiedener Haltungssysteme fur Legehennen hinsichtlich wirtschaftlicher, gesundheitlicher und ethologischer Parameter unter besonderer Berücksichtigung von ausgestalteten Käfigen', *Deutsche tierärztliche Wochenschrift*, 112, pp. 332–342.

Wahlström, A., Tauson, R. and Elwinger, K. (2001) 'Plumage condition and health of aviary-kept hens fed mash or crumbled pellets', *Poultry Science*, 80, pp. 266–271.

Waiblinger, S., Boivin, X., Pedersen, V., Tosi, M.V., Janczak, A.M., Visser, E.K. and Jones, J.B. (2006) 'Assessing the human–animal relationship in farmed species: a critical review', *Applied Animal Behaviour Science*, 101, pp. 185–242.

Wall, H., Tauson, R. and Elwinger, K. (2002) 'Effect of nest design, passages, and hybrid on use of nest and production performance of layers in furnished cages', *Poultry Science*, 81, pp. 333–339.

Wang, G., Ekstrand, C. and Svedberg, J. (1998) 'Wet litter and perches as risk factors for the development of foot pad dermatitis in floor-housed hens', *British Poultry Science*, 39, pp. 191–197.

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.

Warris, P.D., Bevis, E.A. and Brown, S.N. (1990) 'Time spent by broiler chickens in transit to processing plants', *Veterinary Record*, 127, pp. 617–619.

Warris, P.D., Kestin, S.C., Brown, S.N., Knowles, T.G., Wilkins, L.J., Edwards, J.E., Austin, S.D., Nicol, C.J. (1993) 'The depletion of glycogen stores and indices of dehydration in transported broilers', *British Veterinary Journal*, 149, pp. 391–398.

Warris, P.D., Pagazaurtundua, A. and Brown, S.N. (2005) 'Relationship between maximum daily temperature and mortality of broiler chickens during transport and lairage', *British Poultry Science*, 46, pp. 647–651.

Weaver, W.D. and Meijerhof, R. (1991) 'The effect of different levels of relative humidity and air movement on litter conditions, ammonia levels, growth and carcass quality in broiler chickens', *Poultry Science*, pp. 746–755.

Weber, R.M., Nogossek, M., Sander, I., Wandt, B., Neumann, U. and Glünder, G. (2003) 'Untersuchungen zum Gesundheitsstatus von Legehennen in ausgestalteten Käfigen im Vergleich zu Tieren in konventioneller Käfig- und Bodenhaltung', *Tierärztliche Monatsschrift*, 90, pp. 257–266.

Weeks, C.A. and Butterwoth, A. (eds) (2004) *Measuring and Auditing Broiler Welfare*. CABI publishing, Wallingford.

Weeks, C.A. and Kestin, S.C. (1997) 'The effect of leg weakness on the behaviour of broilers', 5th European Symposium on Poultry Welfare, Wageningen, 7 June.

Weeks, C.A., Nicol, C.J., Sherwin, C.M. and Kestin, S.C. (1994) 'Comparison of the behaviour of broiler chickens in indoor and free-range environments', *Animal Welfare*, 3, pp. 179–192.

Weeks, C.A., Danbury, T.D., Davies, H.C., Hunt, P. and Kestin, S.C. (2000) 'The behaviour of broiler chickens and its modification by lameness', *Applied Animal Behaviour Science*, 67, pp. 111–125.

Weeks, C.A., Knowles, T.G., Gordon, R.G., Kerr, A.E., Peyton S.T. and Tilbrook, N.T. (2003) 'New method for objectively assessing lameness in broiler chickens', *Veterinary Record*, 151, pp. 762–764.

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.

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 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*, in press.

Wenzlawowicz, M. von and Holleben, K. von (2001) 'Assessment of stunning effectiveness according to the present knowledge on electrical stunning of poultry in a waterbath', *Archiv für Geflügelkunde*, 65(6), pp. 193–198.

Wilkins L.J, Brown S.N., Phillips A. J. and Warriss P.D. (2003) 'Cleanliness of broilers when they arrive at poultry processing plants', *Veterinary Record*, 153, pp. 701–703.

Wilkins, L.J., Zimmerman, P.H., Leeb, C. and Nicol, C.J. (2004) 'Investigation of palpation as a method of determining the prevalence of keel and furiculum damage in laying hens', *Veterinary Record*, 155(18), pp. 547.

Williams, B., Solomon, S., Waddington, D., Thorp, B. and Farquharson, C. (2000) 'Skeletal development in the meat-type chicken', *British Poultry Science*, 41, pp. 141–149.

Wood-Gush, D.G.M. (1971) *The Behaviour of the Domestic Fowl*. Heinemann Educational Books, London.

Woodward, M. (1999) *Epidemiology: Study Design and Analysis*. Chapman & Hall/CRC, London.

Yu, M.W. and Robinson, F.E. (1992) 'The application of short-term feed restriction to broiler chicken production: a review', *Journal of Applied Poultry Research*, pp. 147–153.

Zhou, W.T., Fujita, M., Yamamoto, S. (1998) 'Effects of food withdrawal and water withdrawal and high temperature exposure on diurnal variation in blood viscosity of broiler chickens', *British Poultry Science*, 39, pp. 156–160.

Zimmerman, P.H. and Koene, P. (1998) 'The effect of frustrative nonreward on vocalizations and behaviour in the laying hen, Gallus gallus domesticus', *Behavioural Processes*, 44, pp. 73–79.

APPENDIX 1

STANDARDISATION OF MANAGEMENT AND HANDLING-BASED MEASURES IN POULTRY

A. Butterworth, S. Waiblinger, B. Algers 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. This chapter outlines the process by which potential management-based measures for poultry were identified. In addition, consideration was given to whether management and handling information may be of value in the advisory aspects of the assessment in which the information produced is 'fed back' to the farmer to help identify risks to animal welfare and causes of poor welfare so that improvement strategies can be implemented.

A large number of potential management measures were identified and it was assumed that not all of these 33 measures will make their way into the Welfare Quality® assessment system. Validity, reliability and feasibility assessments for each measure were not directly assessed. But to assist in any future decision making process a ranking table was produced based on the assessment of the perceived overall value of the measures by the species researchers. A list of questions to assess handling is also presented.

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 assessment protocol. It is also possible that specific management-based measures will prove to be a more feasible alternative to a single or combination of animal-based measures and so for this reason are kept in the final monitoring scheme.

A second aim of a monitoring system can be 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 will not 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.

Thirdly, some management-based measures are proposed 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 less important 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 the 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, the system where the question can be asked is also specified. 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 both broilers and laying hens. The list of questions is presented in Table A1.1.

There are fewer measures recommended to 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. a 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 reliability, however, we have been careful in the wording to avoid 'leading questions' or an implication through the formulation of the question of 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. This information can be used in the training of the assessor.

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 is 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 or 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 further confirmation for this refined list, researchers independently ranked each question on 'how critical it was for welfare' and the average of their score 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 33 measures for poultry 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.

We are aware that there are several measures which are closely related to a resource-based measures e.g. litter is a resource, but how often it is provided and whether it is clean or not is related to management. Some questions relate to animal-based measures, especially those related to health.

Inspection of poultry at regular intervals is considered to be of great importance in terms of observing animals which are diseased or distressed and to ensure that automated systems are working effectively. Additionally, the quantity and especially quality of human-animal interactions affects the birds' level of fear of humans and thus their welfare, by influencing the amount of stress, with potential consequences on health. Thus the actual activities carried out by the stockperson are likely to be important, and it may be of value to assess how the stockperson interacts with the birds by asking questions about the use of specific behaviour and by visual observations of the stockperson's interaction with the birds.

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 suitable 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, mentioned above, related to the manager's use of specific interactions and some observations of handling that we propose to be included, at least during the refining process of the assessment system. 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.

A1.4 MEASURES

The results are presented in the following tables.

Table A1.1 lists the potential measures to assess management and handling practices. The first column gives the number of the question, and the proposed wording of the question is given in the second column. In the third column it is specified whether the question is most relevant to the Welfare Quality® assessment (WQA), to some form of advisory tool (AT) or whether it is mainly of interest for epidemiological study (Epi). The fourth column specifies to what area of measures the question is directed e.g. to hygiene, management of sick animals, loading and slaughter. The fifth column deals with how to assess the measures i.e. whether it is a direct question (Q), whether it is assessed in the building or from paper records (A) or whether it is a combination of question and assessment (Q&A). The sixth column includes a brief description of the approach that the assessor could take to assessing this measure and any evidence, either observed or resulting from questioning, that may be of value. The seventh column is a short description of the reason for including this potential measure. The eight column presents the proposed recording scheme. A simple recording scheme is proposed in most cases where 0= none, 1= some or present but not fully effective, and 2= fully present or effective. The ninth and final column gives the

welfare relevance as scored by the four authors where five was the maximum score possible.

Table A1.2 presents the 33 questions according to their ranking to allow some selection of these measures on the basis of the priority given them by the authors. The rank and overall score are given as well as the number of the question in table 1 and a short version of the actual question.

The final part in the results section deals with the assessment of handling. Table A1.3 presents questions to be directed to the stockman and Table A1.4 lists behaviour to be observed during routine activities of the stockman.

TABLE A1.1 Management and handling practices: listing of potential measures.

N	Description	WQA / AT / EPI	Group	How to assess?	Instruction for assessor	Reason for inclusion	Scoring scheme	Score
1	Is an all in all / out strategy employed for all houses on the site?	AT	Hygiene / cleaning routines	Q	Q= Ask farmer what his system is for stocking the farm	Disease and biosecurity risks are recognised to be higher on farms which allow stepped stocking or introduction of birds of different ages.	0 = No 1 = Yes	3.2
2	After washing, was the house washed with a power wash and disinfected before this flock was placed?	AT	Hygiene / cleaning routines	Q	Q= Ask farmer what his system is for stocking the farm	Protection from disease has been shown to have significant welfare importance	0 = No 1 = Some houses 2 = All	3
3	Are biosecurity measures (barrier, car wash, visitor book) in place.	AT	Hygiene / cleaning routines	Q & A	A= Visible observation of what is actually used: a) on a daily basis b) available but not used. Q= Ask farmer what is available but not used daily.	Protection from disease has been shown to have significant welfare importance	0 = None 1 = Some 2 = Full range	3
4	Is there evidence that the farm implements a policy for using clean overalls and boots for staff (frequent changes) and visitors (new every time)?	AT	Hygiene / cleaning routines	Q & A	A= Visible observation of what is actually used: a) on a daily basis b) available but not used. Q= Ask farmer what is available but not used daily.	Protection from disease has been shown to have significant welfare importance	0 = None 1 = Some 2 = Full range	3.2
5	How many separate houses of broilers or flocks of layers are on this on site at this time?	AT / Epi	Hygiene / cleaning routines	Q & A	Q =Ask farmer what the number of houses on the site is	For farms with a large number of houses there may be increased potential of disease transmission because of movement of people and materials between houses	Number (x)	2
6	Do you have a formal pest control policy	AT	Hygiene / cleaning routines	Q & A	A = Can you find a paper agreement (with a company providing rodent control) that can be checked, or observe traps and measures on the farm Q = Ask the farmer: Do you put traps down yourself? A = Observation - Include also protection from flies - Fly mesh on windows & doors?	Pest (including bird) control may be an effective part of disease control.	0 = No 1 = Yes	2.25

TABLE A1.1 CONT. Management and handling practices: listing of potential measures.

N	Description	WQA / AT / EPI	Group	How to assess?	Instruction for assessor	Reason for inclusion	Scoring scheme	Score
7	Can the farmer adequately describe steps he would take in the case of disease in his birds.	AT / WQA	Management of sick animals	Q & A	Q = Ask the farmer how he deals with, or would deal with a disease outbreak on the farm.	If the flock is affected by disease, the prompt and effective resolution of disease is important for the welfare of the animals. An alternative is a strategy to enable the company or vet to take prompt effective action.	0 = Inadequate response with no evidence of planning 1 = Moderate response indicating able to act to protect birds. 2 = Full response with reference to a health plan, the strategy within company and record based evidence that response is effective	3.5
8	Is there evidence for effective intervention for birds in distress through culling?	AT / WQA	Management of sick animals	Q & A	Q = discuss with farmer actual decision making on culling A = Observe flock and look for evidence of birds which should have been culled.	Prompt, effective and humane culling of diseased or distressed animals will reduce the potential for suffering.	0 = No 1 = Unclear but carries out culling only as last resort 2 Yes – trained personnel, culling to protect animals from prolonged suffering	4.6
9	Do you 'hospitalise' any birds?	AT / WQA	Management of sick animals	Q & A	Q = discuss with farmer actual decision making on use of hospital pen A = Observe flock and look for birds in hospital pen.	Use of hospital pens for poultry is very variable (common in some countries, although influenced by flock size, is rare in broilers). Hospital pens can reflect good welfare (attention to the individual) or poor welfare (a place to put sick animals and then do nothing more, so care in interpretation is required.	0 = Not used 1 Used, but birds kept in hospital pens rather than effective culling policy 2 Regularly and effectively used	4.3
10	Cockerels – fighting between males, and pecking of male birds by more mature females	WQA	Group dymanics	Q & A	A =Observation of pecked animals A =Observation of pecking / fighting behaviour Q =Questioning of farmer about number of birds which have been removed due to pecking / fighting behaviour	Fighting (adult male birds) and pecking (laying hens) are both serious impacts on the birds affected.	0 No males 1 Males, some evidence of aggressive interaction 2 Males present, but no aggressive interaction	3

TABLE A1.1 CONT. Management and handling practices: listing of potential measures.

N	Description	WQA / AT / EPI	Group	How to assess?	Instruction for assessor	Reason for inclusion	Scoring scheme	Score
11	What is the ratio of animals to caretakers (high or low)?	WQA / AT/ Epi	Number of caretakers	Q	Q = Number of stockmen, time available for animal care (rather than tending machines), number of animals. A = Assess the contact time between stockman & animals	Some intensive poultry systems have very large numbers of birds cared for by one person (up to 30,000 birds) whereas some smaller outdoor flocks have low numbers per carer (500 birds). The effective time available for animal care is partly dictated by animal/stock ratios. However, it is clear that automation can improve time available for animal care, but can also require the stockman to become technically proficient at the potential cost to animal care skills.	Number ratio	1: xx or xxx	2.75
12	How much time do you spend each day for observation of animals?	AT/Epi	Number of caretakers	Q & A	Q = Ask about daily routine – time spent walking in the house with the animals. A = Assess the way in which daily tasks are carried out, walk the house with the stockman. Ask about the whole range of tasks which the stockman must carry out.	The time available for observation of each animal may be linked to the quality of animal care. Stockmen have many tasks to perform, some involve contact with the animals (feeding, bedding, dealing with a sick animal) others do not (machine maintenance, administration, tending ground etc.) Time spent observing animals is recognised as of value. Automation may increase time available, but reduce time spent observing animals and create a high dependency on machines to detect environmental or health hazards.	Time a) Total time (minutes) b) Time per animal (minutes)	3

TABLE A1.1 CONT. Management and handling practices: listing of potential measures.

N	Description	WQA / AT / EPI	Group	How to assess?	Instruction for assessor	Reason for inclusion	Scoring scheme	Score
13	Constancy – how many different people care for the poultry?	AT/Epi	Number of caretakers	Q & A	Q = Ask the farmer how many people care for the animals – are there part time staff, staff who cover for holidays etc.	Changes in staffing may lead to alterations in practices and routine – this may, or may not, be detrimental to animal care.	Number (of caretakers)	3
14	Do you like to have contact with the poultry?	AT/Epi	Attitude	Q	Q= Ask the farmer what his favourite tasks are – do they include tasks that involve contact with the animals?	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 via effects on the animals' relationship to humans. See Tables 3 and 4 (below)	1= Yes – Like to have contact 0= No – Do not like to have contact Yes and No – depends upon the reason for the contact	3
15	Are the catching gangs paid hourly rates or piece work rates?	AT/Epi	Loading and slaughter	Q	Q= Are the people who catch the birds paid hourly, or for each job.	If the catching teams are under time pressure and not paid in relation to 'quality' then increased catching injuries and poor handling may result	Salary = Paid hourly Piece rate = Paid for each flock, house or job.	4
16	Have the staff who depopulate the house had training in the care of birds and prevention of injuries during loading and transport?	AT/Epi	Contact with animals Attitude	Q	Q= Have the catching staff had training in ways of reducing damage (and suffering) – for example – have these staff been shown the effects of catching on muscle damage in the carcase.	Training in techniques and a humane attitude to the handling of animals, even if these animals are for slaughter, is considered to be of importance	0= No training 1= Some training 2= Effectively trained	3.5
17	Do you turn the lights up or down when you make the flock walk?	AT/Epi	Management of housing system	Q	A= When the inspector visits the house, what does the farmer do with the lights?	Some evidence suggests that lowering the light levels may make birds more calm – however, low light levels make it very difficult to examine birds adequately. In many countries, houses with wire sides are used and lighting levels are naturally very high – and this does not appear to cause problems. It is abrupt or unexpected light level changes which are usually a concern.	No = Light remain unchanged during flock inspection Yes = Lighting level altered during the inspection	3

TABLE A1.1 CONT. Management and handling practices: listing of potential measures.

N	Description	WQA / AT / EPI	Group	How to assess?	Instruction for assessor	Reason for inclusion	Scoring scheme	Score
18	How many spot checks of the flock in addition to flock walks are made daily?	AT/ EPI/ WQA	Contact with animals Routine health care Attitude	Q	Q= How many times are the animals checked daily – this may be carried out following a routine – i.e. check at intervals prescribed by a plan – or (more usually) animals are checked when time is available. A= It may be possible to observe flock checks going on if there are a number of staff working.	The time available for observation of each animal may be linked to the quality of animal care.	N= Number of checks in addition to flock walks.	3
19	Does a procedure exist for dealing with problems / emergencies?		Problem awareness	Q & A	Q= Ask the farmer what his contingency plan is for fire, power, ventilation, food or water failure, flood.	Emergencies are rare. However, when power fails, the ventilation system fails, or when a fire takes place or when there is flooding, there is huge potential for animal suffering if there are not contingencies in place.	No Yes	3
20	Is there a fire alarm system?	WQA/ AT	Problem awareness	Q & A	A= Observe the fire alarm system Q= Ask when it is tested	Emergencies are rare. However, if a fire takes place there is huge potential for animal suffering if there are not contingencies in place.	No Yes	3
21	Is there a reserve electricity system that powers ventilation and water systems?	WQA/ AT	Management of ventilation/ lighting/ feeding/ biosecurity resources Problem awareness	Q & A	A= Observe the reserve electricity system Q= Ask when it is tested	Power failure is common in some countries.	No Yes	4
22	Do you disinfect the feeder system before the birds are placed in the house?	AT/Epi	Hygiene / cleaning routines	Q	Q= Ask farmer what his system is for disinfecting feeders	Protection from disease through hygiene routines has been shown to have significant welfare importance	0 = No 1 = Some houses 2 = All	3.75

TABLE A1.1 CONT. Management and handling practices: listing of potential measures.

N	Description	WQA / AT / EPI	Group	How to assess?	Instruction for assessor	Reason for inclusion	Scoring scheme	Score
23	Is a policy in place for managing spillage of feed and water?	AT/Epi	Hygiene / cleaning routines Management of ventilation/ lighting/ feeding/ biosecurity resources	Q & A	Q= Ask farmer what his system is for cleaning up spilled feed, and for maintaining drinkers. A= a) Observe – feed on top of ducts and pipes and feed boxes – particularly if old and spoiled b) Observe drinkers and pipework for leaks, and observe litter condition	Spilled feed (inaccessible to birds) is likely to attract pests and may also cause release of mould spores which can lead to increased speed of spoilage of stored feed. Water leaks, particularly from the drinker lines, are a very significant cause of poor litter and so attention to leaking drinkers can protect birds from foot and skin lesions.	0 = No evidence of spilled feed or leaks 1= Some evidence of poor leak / spill maintenance 2= Large amounts of spilled material and / or wet litter below leaking drinkers	3.5
24	If you are producing fertile eggs, how do you feed males and females to control weight gain?	AT/Epi	Mating and genetic management	Q & A	A= Observe feed restriction system for male birds	Feed restriction is a common part of managing male birds in many systems. If well managed, moderate feed restriction can prevent overweight birds, but if poorly managed – it can result in hunger and behavioural effects associated with frustrated feeding activity.	0= No feed restriction 1= Well managed feed restriction which balances the effects of overweight birds, with the feeding needs of the birds 2= Evidence of birds which are suffering long periods without feed	3.5
25	Do you have a routine for maintenance of equipment and for maintenance of litter condition?	WQA/ AT/Epi	Maintenance of equipment	Q	A= Observe conditions in the house – is the litter uniform or are there areas of patchy poor litter. Q= Ask farmer about intermittent problems that he may have with litter and his solutions to these problems.	Poor litter condition may be caused by house and floor conditions, environmental humidity, temperature and ventilation, bird diet and faecal composition, leaking drinkers and poor attention to problems areas within the house. Poor litter affects bird foot and skin health, may affect the ability of birds to dust bathe and lie comfortably, and can result in dirty plumage.	0= Litter is all of good quality 1= Some areas of poor litter quality 2= Poor litter in much of the house	4.25

TABLE A1.1 CONT. Management and handling practices: listing of potential measures.

N	Description	WQA / AT / EPI	Group	How to assess?	Instruction for assessor	Reason for inclusion	Scoring scheme	Score
26	Is slaughter without stunning permitted?	WQA	Slaughter	Q & A	A= Observe system in operation Q= Ask operator about the stunning system used	Scientific understanding now promotes the concept that it is readily possible, and highly preferable, to induce a stunned state in animals before they are killed.	0= None 1= Stunning used but ineffective 2= Effective stunning used	5
27	What diseases is this flock vaccinated against at the hatchery & farm?	AT/Epi	Routine health care	Q	Q= Ask farmer about vaccination programme.	Vaccination against a range of viral (and now some bacterial and parasitic conditions) has been a very significant part in protecting bird health and welfare in many systems. However, it is possible that sometimes vaccination is used when it would be possible to use natural immunity to protect the birds.	Vaccination use should be seen to be appropriate to the real risks presented to the birds.	3
28	Is there a health / welfare plan in place and is it used?	WQA/ AT	Routine health care Attitude Problem awareness	Q	A= Assess the written health and welfare plan. Q= Ask the farmer how he uses the plan.	In many companies, a plan of action to record, track and tackle health and related problems is in place. This health plan can be a very useful tool for the farmer (and the company) but is sometimes seen as a 'paper exercise' rather than a tool of real value.	0= No plan in place 1 = Plan in place, but not used routinely 2= Plan used actively	3.25
29	At what stage would you call a vet when there is a disease problem?	AT/Epi	Problem awareness Routine health care	Q	Q= Ask the farmer to describe some events when he has asked the vet to visit.	In many farm systems for poultry, a vet is available through the company – however, it is common in all systems for the vet to be called at quite a late stage of disease. This is very difficult to assess, but may be measured against the need of animals to receive prompt diagnosis and effective treatment without delay.	The speed and effectiveness of diagnosis and treatment of disease should be noted – but is very difficult to quantify.	3.75

TABLE A1.1 CONT. Management and handling practices: listing of potential measures.

N	Description	WQA / AT / EPI	Group	How to assess?	Instruction for assessor	Reason for inclusion	Scoring scheme	Score
30	Any environmental enrichment? E.g. bales of straw, perches, toys	WQA/ AT	Enrichment	Q & A	A= Assess any available enrichments – toys, brassicas to peck at, straw bales, clean straw for scratching in, dust or sand bathing areas. Q= Ask farmer about enrichments used for times that are not assessed – i.e. for chicks etc.	Enrichment in intensive houses may take the form of – toys, brassicas to peck at, straw bales, clean straw for scratching in, dust or sand bathing areas. For birds with outdoor access or free range birds, there may be no additional enrichments provided. Many agree birds can use appropriate enrichment objects or resources (e.g. straw).	0= No enrichment 1= Some enrichment 2= Well managed and diverse enrichment.	3.5
31	Are your birds beak trimmed?	WQA,/AT	Mutilations	A	A= Assess whether birds beak trimmed, and how well they are beak trimmed.	Beak trimming is common for laying birds, but the need for trimming can be reduced by management of laying birds. Some birds may be poorly beak trimmed leading to neuroma formation and potential discomfort.	0=Not beak trimmed 1= Beak trimmed, all adequate length 2= Some excessively short beak trimmed	4
32	What is the floor made of? Does the floor incorporate a damp proof membrane?	AT/Epi	Management of housing system	A	A= Scrape away the litter and assess whether the floor is concrete, packed earth or packed rock.	Northern countries, the floor is usually concrete; southern countries, it is common to have earth of compacted stone under the litter. Control of moisture in the litter may be more difficult in non-concrete floors – but many poultry keepers manage non-concrete floors well. Floors not made of concrete are very difficult to fully clean between flocks.	1=Concrete floor 2= Earth or packed stone floor	3.5
33	Are any computerised/ time clock systems used to control lighting?	AT	Management of ventilation/ lighting/ feeding/ biosecurity resources	Q & A	Q= Ask farmer to explain how the lighting is controlled.	Lighting patterns can provide variable day length at different times in the birds' life.	0=No automated lighting system 2 1= Fully automated lighting system	

TABLE A1.2 Rankings of the 33 potential measures listed in Table A.1.

Rank	Score	Table number	Description
1	5	26	Is slaughter without stunning permitted?
2	4.6	8	Is there evidence for effective intervention for birds in distress through culling?
3	4.3	9	Do you 'hospitalise' any birds?
4	4.25	25	Do you have a routine for maintenance of equipment and for maintenance of litter condition?
5	4	15	Are the catching gangs paid hourly rates or piece work rates?
5	4	21	Is there a reserve electricity system that powers ventilation and water systems?
5	4	31	Are your birds beak trimmed?
6	3.75	22	Did you disinfect the feeder system before the birds are placed in the house?
6	3.75	29	At what stage would you call a vet when there is a disease problem?
7	3.5	7	Can the farmer adequately describe steps he would take in the case of disease in his birds.
7	3.5	16	Have the staff who depopulate the house had training in the care of birds and prevention of injuries during loading and transport?
7	3.5	23	Is a policy in place for managing spillage of feed and water?
7	3.5	24	If you are producing fertile eggs, how do you feed males and females to control weight gain?
7	3.5	30	Any environmental enrichment? E.g. bales of straw, perches, toys
7	3.5	32	What is the floor made of? Does the floor incorporate a damp proof membrane?
8	3.25	28	Is there a health / welfare plan in place and is it used?
9	3.2	1	Is an all in all / out strategy employed for all houses on the site?
9	3.2	4	Is there evidence that the farm implements a policy for using clean overalls and boots for staff (frequent changes) and visitors (new every time)?
10	3	2	After washing, was the house washed with a power wash and disinfected before this flock was placed?
10	3	3	Are biosecurity measures (barrier, car wash, visitor book) in place.
10	3	10	Cockerels – fighting between males, and pecking of male birds by more mature females
10	3	12	How much time do you spend each day for observation of animals?
10	3	13	Constancy – how many different people care for the poultry?
10	3	14	Do you like to have contact with the poultry?
10	3	17	Do you turn the lights up or down when you make the flock walk?
10	3	18	How many spot checks of the flock in addition to flock walks are made daily?
10	3	19	Does a procedure exist for dealing with problems / emergencies?
10	3	20	Is there a fire alarm system?
10	3	27	What diseases is this flock vaccinated against at the hatchery & farm?
11	2.75	11	What is the ratio of animals to caretakers (high or low)?
12	2.25	6	Do you have a formal pest control policy
13	2	5	How many separate houses of broilers or flocks of layers are on this on site at this time?
13	2	33	Are any computerised/ time clock systems used to control lighting?

TABLE A1.3 Stockman–animal interaction questionnaire.

How often do you								
Talk to birds when approaching them	never	1	2	3	4	5	6	7 always
Talk to the birds when walking through the group	never	1	2	3	4	5	6	7 always
Do you knock on the door or give some other signal to make birds aware before entering the animal house?	never	1	2	3	4	5	6	7 always
Are birds regularly caught (for weighing)?	never	1	2	3	4	5	6	7 regularly
Wave with the hands to move them	never	1	2	3	4	5	6	7 always
Push birds with the feet to move them	never	1	2	3	4	5	6	7 always
Shout at birds to make them move	never	1	2	3	4	5	6	7 always
Use a stick to move birds by herding them	never	1	2	3	4	5	6	7 always
Are whole grains fed (in the litter, in the feeder or out on range)	never	1	2	3	4	5	6	7 always
Are activities which involve contact with birds carried out in a routine fashion or simply whenever time permits during the day?	never	1	2	3	4	5	6	7 always

TABLE A1.4 Observation of interaction during routine activity.

Interaction		
Gentle vocal interactions (softly talking to animals)	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 feet with low force, dominant talking to animals to make them stop behaviours)	Never occurred 1 _____	10 often occurred
Carefulness/suddenness of movements	Careful always 1 _____	10 often makes sudden movements

APPENDIX 2

STANDARDISATION OF RESOURCE-BASED MEASURES IN POULTRY

A. Butterworth and E. Fröhlich

A2.1 SUMMARY

The first and main role of resource-based measures in Welfare Quality® is to address issues that can not be dealt with using animal-based measures. Resource-based measures refer to measures taken in the environment in which the animal is kept. The objective of this review is to select appropriate resource-based measures and to compile work instructions to give clear guidelines regarding the accurate and reliable collection of these measures. Validity, reliability and feasibility assessment of each measure were not directly assessed. But to assist in any future decision making process, tables of potential resource measures are provided, along with suggested units, and a ranking of the potential 'value' of each measure.

A2.2 INTRODUCTION

Resources such as housing system, ventilation system, feeding, lighting, temperature control, nest box design, bedding and litter quality as well as the spatial distribution and availability of these resources during the birds' life significantly affect bird welfare. Existing 'records' e.g. farm records kept by producers and companies, and the immediate 'record' available by questioning stockmen or animal caretakers on the use of resources may provide useful information.

The main role of resource-based measures in Welfare Quality® is to address issues that can not be dealt with using animal-based measures and so they may represent the only option for a particular dimension of animal welfare to be included in the final monitoring system. It is also possible that specific measures will prove to be a more feasible alternative

to one or a combination of animal-based measure and so are for this reason kept in the final monitoring scheme instead of the animal-based measure.

Secondly, questions are included which are primarily designed to give feedback to the farmer/manager on what factors might have contributed to results from some of the animal-based measures. For example a high level of ammonia in the air in a poultry house would lead an advisor to look at the farmers response to questions related to ventilation and litter management and routines for maintaining the ventilation system.

Thirdly, some measures are proposed for scientific reasons since the data gathered presents a unique opportunity for epidemiological analyses aimed at furthering our knowledge of animal welfare. It must be accepted that this is a secondary aim of the work in this deliverable and so if it is found there is a shortage of time to ask these questions, then these are the questions that should be excluded first. However, we include them since they could be a valuable scientific spin off of Welfare Quality®.

A2.3 APPROACH TAKEN

Although animal-based measures are the measure or choice in Welfare Quality®, since they are independent of housing system, this is not the case for resources 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.

Once a first list of the questions was developed each researcher went through the list of questions and ranked them according to the following questions:

1. How critical the issue addressed in the question was for animal welfare (low to high);
2. The portion of life referred to in the question (short to long period of the animal's life);
3. The time needed to gather the data (quick or time consuming);
4. The quality/reliability of the data (poor to good);
5. Whether there was a critical window during which the measurement must be taken (yes to no it could be gathered anytime).

Tables were used to help understand the strengths and weaknesses of the possible resource measures, the advantages / disadvantages were listed and then visual analogue lines used to score the merits of each using the five criteria above. Although other criteria could have been selected, these appeared to give a simple method for ranking the measures and to help in the decision making. Higher scores indicate a better mix of welfare significance, practicality, data collection ease and specificity and the significance of critical windows for assessing the specific measure.

A2.4 MEASURES

Table A2.1 includes the lists of potential resource measures, along with short methodologies for collection of data, suggested units, and the potential 'value' of each measure. Critical control points (CCP) are given where appropriate. As a further aid to development of the final protocol, we have listed whenever we think one of our measures relates to another measure that might be included in the Welfare Quality® assessment protocol. A large number of potential measures is provided and it is clear that there will need to be selection from this large collection of possible measures before inclusion in any practical welfare assessment scheme.

The members of the group were chosen for their experience in this area so there had already been a large amount of filtering of questions and most questions ranked favourably in these areas and researchers used the results of the ranking to further refine the list. In Table A2.2 the measures are ranked and this is intended to help during the development of operational protocols if there needs to be elimination of some questions. 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.

TABLE A2.1 Comparison of different resource measures for broilers and laying hens.

No.	Criteria	Advantage	Disadvantage	Score	Links to other measures in WQ
1	Space (surface cm2/animal) Space on a simple flat floor may be readily calculated - however, in environments with perches, feeders, additional tiers, boxes etc. useable space is complex to calculate	Increased space may allow animals to show an increased range of behaviours, incl. avoidance behaviours, and may permit choices of local environment to be made within a house environment. High stocking leads to pressure on air, ventilation and food / water resources	Limited space may reduce all of the factors mentioned but may (in cold conditions) improve the thermal environment of the house	4.1	Behaviours Resource access (food, water) Disease
2	Feeders - General CCP a) Poorly placed b) Technical failure c) Not filled sufficient times daily d) Not sufficiently filled	Nutritionally balanced, healthy feed and proper storage without spoilage or contamination (silo etc.) are beneficial. Absence of hunger is the result if feed is provided at all times in sufficient amount and easy accessible	Unbalanced feed may cause malnutrition, poor hygiene and storage of feed may destroy feed value and cause disease	4.5	Aggression Disturbance Hunger Body weight gain Laying performance, Lameness Feed composition Feed storage
3	Chain feeder (length of chain feeder cm/animal) CCP – (in addition to general) a) Insufficient length b) Accessibility	Fresh food can be available (fed) many times a day. Close monitoring of feed intake, Food Conversion (FC) and Food Conversion Ratio (FCR) calculation possible which are of commercial importance	Unequal distribution of feed. The equipment can form a physical barrier across the house, especially when birds are feeding heavily (crowding along the feed rail).	4.65	Aggression Disturbance Hunger Body weight gain Laying performance Lameness Feed composition Feed storage
4	Pan feeder (length of circumference cm/animal) CCP – (in addition to general) a) Insufficient length b) Accessibility	Pan feeders can be used as a 'shelter' resource for some birds. Equal distribution of feed. Close monitoring of feed intake, FC and FCR possible	One pan can be dominated by one bird. Height decisions for feeders can affect the birds use of the feeders. Pan design is critical to the use of the pan to prevent dominance & 'conquerance' behaviour.	4.65	Aggression Disturbance Hunger Body weight gain Laying performance Lameness Feed composition Feed storage
5	Manual feeding (hopper) CCP – (in addition to general) a) Insufficient length b) Poor storage	Stockman spends more time in close contact with the animals.	Time taken to feed manually (time consuming) may restrict the time available for other activities. Close monitoring of feed intake, FC and FCR is restricted	4.2	Aggression Disturbance Hunger Body weight gain Laying performance Lameness Feed composition Feed storage

TABLE A2.1 CONT. Comparison of different resource measures for broilers and laying hens.

No.	Criteria	Advantage	Disadvantage	Score	Links to other measures in WQ
6	Drinkers – General (number of birds per drinker) CCP a) Insufficient number or space b) Poorly placed c) Poor water supply d) Poorly maintained	Good quality water (bacteriology) and freshness are important. Absence of thirst if water is provided at all times in sufficient amount and with easy accessibility.	Contamination, poor hygiene and storage of water may cause disease	4.2	Aggression Disturbance Thirst Body weight gain Laying performance Litter quality Ammonia,
7	Nipple drinkers (number of nipples/animal) CCP (in addition to general) a) Spillage on litter, belts & manure pits b) Poorly cleaned pipework	Delivery of clean water straight to the animal. a) with drip cups or b) without drip cups If water leaks the litter and manure are protected.	In hot temperatures may not deliver water fast enough. Without drip cups: Leaky drinkers lead to poor litter quality or if placed above manure belts (pits) to wet manure. Both increases ammonia production	4.8	If tubes are poorly cleaned can lead to increased disease risk
8	Cup drinkers (nipple feeds cup) (number of animals/cup) CCP (in addition to general) a) Too many, can lead to poor hygiene within the cups b) Poorly cleaned cups can lead to disease	Bird can drink more easily, more at a time and in normal posture/way (bigger volume available).	Hygiene – contamination of the cups, particulary if close to end of the line. Some cups may not work reliably, either they overfill or remain empty.	4.5	If cup are poorly cleaned can lead to increased disease risk
9	Bell drinkers (length of circumference cm/animal) CCP (in addition to general) a) Spillage on litter b) Poorly cleaned gutter can lead to disease	Bird can drink more easily, more at a time and in normal posture/way (bigger volume available).	Hygiene – contamination of the bell. Easily sway and water spills over causing litter moisture problems	4.4	If gutter poorly cleaned can lead to increased disease risk
10	Nest boxes – General CCP a) Number of nests available b) High number of mislaid eggs c) Dirty nests (Hygiene) d) Number of devaluated eggs e) Aggression in front of nests	Use of nests promotes: Normal and undisturbed laying behaviour, the need of hens to separate for laying, calm daily laying period. Reduces: Pre-laying stress and frustration Number of mislaid eggs, and thus labour costs and poor egg quality.	Economical aspects (costly) Must be managed	4.25	Laying performance Aggression Time budget (daily laying period) Vocalisation Pre laying behaviours 'Conquerance' for nests

TABLE A2.1 CONT. Comparison of different resource measures for broilers and laying hens.

No.	Criteria	Advantage	Disadvantage	Score	Links to other measures in WQ
11	Single nest (number of animals / nest) CCP (in addition to general) a) More than one bird per box b) Poor nest accessibility – what is in front of the nest	Separation – non-conquerance / bullying behaviour	Single bird in open fronted box may be target of aggression	4.25	
12	Group nest (number of animals / nest) CCP (in addition to general) a) Size of the box b) Depth of the box c) Poor nest accessibility – what is in front of the nest d) Temperature within group nests e) Technical failure of closing door systems	Social imitation / facilitation is possible. Technically (and economically) advantageous to manage group nests. May be more protected by curtains (retreat behaviour possible)	Can be a disadvantage if the nest size is too large Mortality by crushing in closing door systems and if the nest box is too deep Thermal behaviours in overcrowded box, if too large the birds also carry out other behaviours in the box.	4.25	Thermal behaviours Hygiene – (dirty nests) Injury Mortality (by crushing)
13	Nest floor (material) CCP (in addition to general) a) Plastic / rubber mat b) Litter c) Artificial nest d) Wire mesh/ coated wire mesh	a) Plastic mat / rubber mat – Comfort, Hygiene b) Litter – Comfort, Can show nesting behaviour c) Artificial nest – comfort , hygiene d) Wire mesh / coated wire – hygiene	a) Plastic mat / rubber mat – Non natural material so bird unable to 'build' nest – so restricts nesting behaviour b) Litter – can have poor hygiene leading to feather contamination and skin disease c) Artificial nest – sometimes more than 1 bird tries to use nest simultaneously d) Wire mesh / coated wire – comfort damage to plumage, feet & nesting behaviour restricted	4.35	Thermal behaviours Hygiene – (dirty nests) Injury
14	Perches (length of chain feeder cm/animal) CCP a) Shape b) Material c) Accessibility, placement d) Are there perches e) Placement of perches f) Rest period during daytime present g) Disturbance before rest at night h) (Breast) injuries	Allow perching behaviour Allow use of the foot to grip Perching leads to better quality of rest Raised perches may be beneficial to birds rest, preening, retreat and withdrawal from others, anti-predator behaviour and may improve physical fitness and use of space	Collisions with perches (physical arrangement of perches) or physical damage caused by slipping on perches Shape may provoke foot pad problems	3.5	Resting behaviour Disturbance Injuries Behaviour (prolonged hesitation to climb down from poorly placed perch systems) Time budget (daily resting period)

TABLE A2.1 CONT. Comparison of different resource measures for broilers and laying hens.

No.	Criteria	Advantage	Disadvantage	Score	Links to other measures in WQ
15	Flooring – General	See below for flooring types	See below for specific flooring issues	3.8	Body injuries (injuries to toes)
	CCP				Foot pad problems
	a) Not in good state of repair				Trapping injuries
	b) Weak construction (relevant to raised floors supporting wire mesh or slats)				
16	Fully perforated floor	Faeces separation	Poor rest if no foot closure possible	3.7	Poor quality of rest if cannot close the feet, increased potential for injuries
	(plastic / wooden slats)	Plastic floors are likely to be more foot friendly than wire mesh	Animals get trapped or injured if not in a good state of repair		
	CCP (in addition to general)				
	a) Dimension of slats not strong enough				
	b) Does not allow automatic foot closure				
	c) Gaps between floor elements				
17	Wire mesh	Faeces separation	Foot damage	3.9	Increased potential for injuries
	(wire mesh – dimension of aperture cm^2)		Animals get trapped or injured if not in a good state of repair		
	CCP (in addition to general)				
	a) Dimension of wire not strong enough				
	b) Bowing, fixation of the wire mesh elements				
	c) Gaps between floor elements				
18	Solid floor – covered with litter	May shorten the claws in layers	Hygienic risk	4.0	Uncovered solid floor creates an increased disease risk
	(concrete)		Comfort		Injury through slipping / abrasion
	CCP (in addition to general)		Thermal comfort		
	a) Not covered with at least some litter material				
	b) Too abrasive nature of concrete				
19	Litter – General	Promotes:	Must be well managed	3.7	Behaviour
	(litter surface cm^2/animal)	Lying comfort, foot & thermal comfort	Must be appropriate depth (hygiene)		Mortality (due to cannibalism)
	CCP	Occupational enrichment	Litter must not be wet, greasy or capped		Disease (coccidiosis, worms)
	a) No litter material	Dust bathing & foraging	(hygiene)		Time budget (daily dust bath / f oraging period)
	b) Bad litter quality	Reduces:	If too deep / wet etc. is a disease hazard		
	c) Poor litter management	Behavioural problems (feather pecking, cannibalism, sham dust-bathing etc.)			

TABLE A2.1 CONT. Comparison of different resource measures for broilers and laying hens.

No.	Criteria	Advantage	Disadvantage	Score	Links to other measures in WQ
20	Litter material (whole straw, chopped straw, wood shavings, sawdust, sand, rice husks, paper, peat, earth) CCP (in addition to general) a) Whole straw b) Chopped straw c) Wood shavings d) Mixed wood shavings/straw e) Sand f) Peat / earth	a), b), d) Good occupational material for layers c) Good occupational material for chicks e) Good material for dust-bathing	a) Poor litter material for broilers e) Poor absorption of humidity e) Poor occupational material	4.8	More foraging/pecking (in whole straw and mixtures) Feather pecking Foraging dust bathing
21	Litterquality (uniformity/texture/friability/humidity) CCP (in addition to general) a) Wet b) Not friable c) Capped d) Greasy e) Too shallow f) Too deep g) Too uniform, may discourage foraging h) Too dry = increased dust i) Non uniform area (wet areas) can affect birds in these areas	Dry litter is a comfortable and hygienic bedding material (broilers, layers) Uniformly dry/good litter means that there are no 'risk areas' for the birds Good quality litter allows – dust bathing, scratching etc For layers, working the litter (scratching, foraging) is an important occupational activity	Wet, greasy or capped litter is an unhygienic and thermal/wet uncomfortable material Many houses have areas of poor litter within an overall dry litter – leading to local problems for the birds in these areas Poor quality litter does not permit these behaviours and has bird cleanliness and comfort implications Very dry litter is a respiratory hazard (humans!)	3.6	Plumage condition Behaviour (if litter unavailable for dust bathing/scratching) Feather peckingC Cannibalism a.s.o Feather pecking (if substrate too uniform) Foraging (reduced if too uniform, wrong material) Dust bathing (if too wet deep etc.)
22	Management CCP (in addition to general cycle in some systems) a) Stockman needs to monitor litter for wet areas (remove, improve) b) Addition of material during	If well managed = advantage to the bird	If poorly managed = disadvantage to bird (hygiene, comfort, skin pathologies, behaviour)	3.6	Behaviour Disease Foraging Plumage condition Feather pecking (all affected by litter management)
23	Light intensity (lux) CCP a) Long term adequate level (very low levels of light are a welfare concern) b) If very low light levels are being used, this may mean that this is masking other problems	High light levels may allow high orientation and visually based behaviours of birds, and also allows good observation of stock (Is 'lux' a good measure, or is the sensitivity of birds to different wavelengths more critical – some debate about this)	Risk of cannibalism (layers) at high light levels	3.2	Behaviours Aggression Cannibalism Disease (ocular abnormalities – blindness/cataracts)

TABLE A2.1 CONT. Comparison of different resource measures for broilers and laying hens.

No.	Criteria	Advantage	Disadvantage	Score	Links to other measures in WQ
24	Light patterns CCP a) Length of light periods b) Length of dark period c) Dormant task periods	Having light dark/periods provides resting/recovery periods Laying hens become stressed if they do not have dark periods Sufficient dark periods can increase activity levels during the light period (broiler) Dormant task periods gives time to carry out these tasks	Laying hens – advantageous to have dark periods as they become very restless without periods of sleep Very long light periods may be considered as very unnatural Return to feeders after dark period considered a possible hazard Broiler – Potential risk of piling up/panic during the dark period if disturbed	4.75	Behaviour Feeding behaviour Resting behaviour Vocalisations if insufficient rest period (particularly layers) (frustration)
25	Light spectral (frequency Hz) CCP a) Low frequency fluorescent lamps (<70Hz) can be seen as flicker by birds b) High frequency fluorescent lamps (>70Hz) c) If using a single frequency source (LEDs) may get one frequency band effects		May be able to detect flicker Single frequency source lights may cause ocular damage	4.4	Nervous behaviour resulting from flicker
26	Natural daylight CCP a) Windows present b) Number, placement of windows c) 'Spots' of bright daylight can cause problems	Whole spectrum including UV Hygiene implication of UV light (biocidal activity of UV) May be aesthetic considerations (perception of public)	High light intensity may lead to cannibalism (layers)	3.2	Piling up or smothering behaviour is possible (if bright patches of light)
27	Ventilation CCP a) Present? b) Working? c) Effective? d) Source of clean air at animal level? e) Acceptable levels of dust/gases?	Essential to maintain good air quality	Good air quality/ventilation requires investment (in money & management)	2.75	Disease Behaviour Activity Thermal coping behaviours

TABLE A2.1 CONT. Comparison of different resource measures for broilers and laying hens.

No.	Criteria	Advantage	Disadvantage	Score	Links to other measures in WQ
28	Biosecurity resources CCP a) Boot washes? b) Personal hygiene? c) Monitoring visitors? d) Disinfection/preparation of house? e) Disinfection of vehicles?	Hygiene & disease control	Biosecurity may limit/restrict the contact between humans and animals and restrict the use of outdoor runs	3.95	Disease Human–animal interactions
29	Alarm systems For: a) House/environment control b) For fire etc. c) Back up generators CCP a) Complicated systems may be in place, but may not be effective – how often is real testing carried out b) Automated control systems should be tested that they actually function	Can protect animal life and quality of life in the case of unexpected failure of, for example, ventilation system – and protect	Automated systems may reduce human sensitivity to animal responses	4	Injury Disease Flock mortality disaster
30	Manure removal a) Dropping pits b) Manure belts c) Partial removal of litter CCP a) Ammonia levels (may be monitored subjectively by inspector reaction to air (>5ppm detect, >20 cough, >70 v. unpleasant) –may be possible to measure using dragger tubes	Deep litter (if well managed) contributes to house temperature and can be a self sustaining bedding material	Air quality considerations if litter not removed (for layers) Rodent problems if litter not removed (for layers) Poorly managed deep litter (broilers) may cause feather soiling, foot and hock lesions, breast & skin conditions. (Management of litter critical)	3.3	Disease (Air quality – increased ammonia, dust, etc.)

TABLE A2.1 CONT. Comparison of different resource measures for broilers and laying hens.

No.	Criteria	Advantage	Disadvantage	Score	Links to other measures in WQ
31	Maintenance of resources CCP a) Do you see evidence of poor maintenance, construction of the building, etc.? b) Some 'modifications' to mechanical systems may be real attempts to improve situation, but implications should be understood c) If resources are poorly designed for human maintenance then they may be poorly maintained. Any evidence of this (difficult to maintain systems?)	Well maintained resources may be beneficial to welfare	Poorly maintained resources may be a risk of injury & disease and a hazard to welfare	3	Disease Injury Human–animal interaction Hunger Thirst
32	Pest (rodent) control CCP a) How much evidence of rodent activity seen? b) If bait is used, it should not present a risk to the birds	Good rodent control reduces disease & animal disturbance (layers)	Poor rodent control may lead to a disease and disturbance risk (layers)	3	Behaviours (disturbance of birds) Disease
33	Parasites (mites, flies, litter beetles) CCP Evidence of mite infestation is a welfare/health risk in layers	Good parasite control reduces disease & animal disturbance	Mites may cause anaemia and irritation Beetles may carry disease and alter the litter properties Flies are an irritant and may carry disease	2.95	
34	Predators (birds, carnivores) CCP a) Are there protection measures in place? b) Provision of cover / field shelters if birds of prey are present?	Good predator control and prevention reduces disease & animal disturbance	Predation, fear, injury	2.35	Mortality Fear Behaviours

TABLE A2.1 CONT. Comparison of different resource measures for broilers and laying hens.

No.	Criteria	Advantage	Disadvantage	Score	Links to other measures in WQ
35	***Placement of resource & accessibility of resource CCP a) Can animals reach the resource without difficulty or without challenge from other animals? b) Is there a structure/design of the environment to minimise disturbance between animals performing different activities? c) Has spacing of animals been considered to avoid pecking/aggressive interactions? d) If equipment (resources) are over complicated, then people will not use them.	Resources must be appropriately placed The technical specifications for many resources and their location/position are precise details which we consider to be required only if the specific resource measure is considered		3.3	Behaviour (Aggression, if resource difficult to access)

TABLE A2.2 Ranking of the proposed measures.

Rank	No.	Measure	[A]Score
1	20	2. Litter material:	4.8
1	15	Flooring 1. General:	4.8
2	7	1. Nipple drinkers	4.75
3	4	2. Pan feeder	4.65
3	24	Light patterns	4.65
4	8	2. Cup drinkers	4.5
4	3	1. Chain feeder	4.5
5	25	Light spectral Frequency	4.4
5	2	Feeder 1. General:	4.4
6	9	3. Bell drinkers	4.35
7	11	1. Single nest	4.25
7	12	2. Group nest	4.25
7	13	3. Nest floor	4.25
8	6	Drinkers 1. General:	4.2
8	10	Nest boxes 1. General:	4.2
9	1	Space	4.1
10	29	Alarm systems	4
10	5	3. Manual feeding	4
11	18	3. Solid floor – covered with litter (concrete)	3.95
12	28	Biosecurity resources	3.9
13	17	2. Wire mesh	3.8
14	16	1. Fully perforated floor	3.7
14	19	Litter 1. General:	3.7
15	21	3. Litter quality	3.6
15	22	4. Management	3.6
16	14	Perches	3.5
17	30	Manure removal	3.3
17	35	Placement of resource	3.3
18	23	Light intensity	3.2
18	26	Natural daylight	3.2
19	31	Maintenance of resources	3
19	32	Pest (rodent) control	3
20	33	Parasites	2.95
21	27	Ventilation	2.75
22	34	Predators (birds, carnivores)	2.35

Notes: The nos. of the measures in the second column refer to the numbers in the first column in Table A2.1.

APPENDIX 3

SAMPLING AND STATISTICS: NOTE ON SAMPLE SIZE REQUIRED FOR PROPER INFERENCE ABOUT PROPORTIONS

B. Engel and K. van Reenen

A3.1 INTRODUCTION

Suppose that we are interested in the proportion of defective individuals, i.e. animals with a reduced welfare, as a reflection of quality of welfare due to management at a farm. Assume that we are able to collect a (quasi) random sample of animals of size n at the farm and that x is the number of defectives in the sample. Also suppose that our judgement of an animal is faultless; there are no false positives or false negatives.

The proportion x/n of defectives in the sample is the *apparent* proportion of defectives. Obviously, unless we sampled all animals, this is not the *true* (unknown) proportion p at the farm, due to sampling variation. For the moment, let us assume that the number of animals at the farm is very large. This conveniently allows us to postpone the issue of infinite or finite population size to a later moment.

Suppose that 0.10 is a critical threshold: a true proportion p below 0.10 is acceptable, but a true proportion p above 0.10 is not. If we warn a farmer that the proportion of defective animals at his farm is too large, we want this warning to be correct. Incorrect warnings would make the farmer lose faith in the monitoring system and in due course commitment would be lost. Usually, we cannot be completely sure that a warning is correct. But we can require the probability for the warning to be incorrect to be small, say smaller than 0.05.

Suppose that we give the farmer a warning when x is larger than some boundary value b. So, $x = b$ defectives in the sample is just acceptable, but $x > b$ is not. Formally stated, we want the probability for x to exceed b to be smaller than 0.05 when the true proportion p is below 0.10. When we call the latter probability α, we require:

$$\alpha = \text{Prob}(x > b, \text{ when } p \leq 0.10) \leq 0.05$$

The boundary value *b* will depend on the sample size n: the larger n, the closer we can choose *b*/*n* to the critical level of 0.10. How large do we want *n* to be? To answer this question, we need to decide what values for *p* larger than 0.10 we want to be able to detect with some confidence. Ideally we would like to detect even the smallest departures, such as p = 0.101. But that is not possible in practice. The reason is that the data generated with p = 0.099 (an admissible value) are virtually the same as the data generated with p = 0.101 (an inadmissible value). Only for a very large sample size n we would be able to see the difference between the two true proportions that are close together, but on different sides of the critical threshold. So, we need to be more modest. Let us say that at least we want to be able to detect a value of *p* as large as 0.2 with a probability larger than or equal to 0.95. Formally:

$$\beta = \text{Prob}(x > b, \text{ when } p = 0.2) \geq 0.95$$

The idea of a small probability α for a false warning and a large probability β to detect a certain departure from the critical threshold (the maximum allowable value for the true proportion of defectives) fits quite natural into the frame work of statistical hypothesis testing. The advantage of that is that we can use the well developed theory of hypothesis testing to determine the sample size n and associated boundary value b. Formally, we test the hypothesis:

$$p \leq 0.10$$

When we reject this hypothesis, because the number of defectives *x* in the sample is too large (when *x* exceeds an appropriately chosen boundary value *b*), we give a warning to the farmer. The first probability of falsely accusing a farmer is called the size α of the test. The second probability β to detect the difference between a true proportion $p = 0.2$ and the critical threshold 0.1, i.e. the probability to reject the hypothesis $p < 0.10$, is called the power of the test. We can determine the sample size n for a size $\alpha \leq 0.05$ test of the hypothesis $p < 0.10$, such that the power β to detect $p = 0.2$ (the alternative hypothesis in statistical jargon) is at least 0.95.

A3.2 SAMPLE SIZES REQUIRED FOR LARGE FARMS

In the table below, the required sample size *n*, and associated boundary value b, to detect the difference between the true proportion $p = 0.2$ and the critical threshold 0.1 are reproduced ($\alpha \leq 0.05$ and $\beta \geq 0.95$). The required sample size n can be quite large. When we are more lenient, for instance increase the true proportion to $p = 0.3$, a smaller sample size n will be required. Therefore, sample sizes for values $p = 0.2, 0.25 \ldots 0.50$ are listed.

The sample sizes will depend on the critical threshold (e.g. 0.1). Results are also presented for a larger critical threshold of 0.2 (Table A3.1).

A3.3 EXAMPLE

Suppose that the critical threshold is 0.1, i.e. we regard a true proportion p that exceeds 0.1 as inadmissible. We want to be able to warn a farmer when the true proportion is as large as $p = 0.3$. With sample size $n = 41$ we will be able to detect such a departure from 0.1 with a probability of at least 0.95. With a sample size of $n = 41$, we reject the hypothesis $p < 0.10$, and consequently give a warning to the farmer, when the number of defectives x in the sample is more than 7 (= b), i.e. when the apparent proportion x/n in the sample is larger than 0.171 (= b/n).

So, with $n = 41$, we warn the farmer when $x = 8$ or more, but we do not warn the farmer when $x = 7$ (or less). We cannot warn the farmer when $x = 7$, despite the fact that the apparent proportion $x/n = 7 / 41 = 0.171$ exceeds the critical threshold of 0.1, because in that case the probability of a false warning would exceed 0.05. In other words, when the true proportion is $p = 0.10$ there is still a probability of finding x as large as 7 that exceeds 0.05. Only when x is as large as 8, we can decide that the probability for this to occur for $p \leq 0.10$ is too small (below 0.05), and we reject the hypothesis $p \leq 0.10$ in favour of $p > 0.10$ and warn the farmer.

TABLE A3.1 Required sample size n (infinite population size, i.e. large farms).

True proportion p	Sample size n	Boundary b: $x > b$ means hypothesis rejected and farmer warned	Proportion b/n: when apparent proportion > b/n, hypothesis is rejected and farmer is warned
Critical threshold = .1*			
.20	135	19	.141
.25	70	11	.157
.30	41	7	.171
.35	31	6	.194
.40	24	5	.208
.50	13	3	.231
Larger critical threshold = .2			
.25	756	169	.224
.30	204	50	.245
.35	98	26	.265
.40	60	17	.283
.45	40	12	.300
.50	28	9	.321

Notes: * above this level the true proportion is considered to be unacceptably large). Hypothesis $p \leq .1$ is tested and rejected when $x > b$. Size $\alpha \leq 0.05$ and power $\beta \geq 0.95$.

A3.4 TAKING FINITE HERD SIZE INTO ACCOUNT

The aforementioned sample sizes are derived assuming an infinite population size. This means that the results apply to relatively large farms, chicken farms for instance, where population sizes can be expected to be in the order of thousands if not ten thousands. For dairy cows however, farm size can be as low as 50 cows, while a relatively large farm may have 200 cows. What about these farms of relatively small size?

Suppose that a farm has a herd of 50 animals. In that case we could inspect all animals, i.e. $n = 50$. The sample and the population (the herd) would coincide and the apparent proportion x/n and the true proportion p would be the same. In that case we could choose $b = 5$, which is the maximum number of allowable defective cows when $p = 0.10$, and give a warning when the number of defectives . Because we know the exact value of p, we are in the ideal situation that $\alpha = 0$ and $\beta = 1$.

So, by virtue of a small herd size, we can detect any value of p that exceeds 0.10, without the risk of giving a false warning, when we sample all animals. Actually, to reach a required power $\beta = 0.95$ for $p = 0.2$, we do not need to sample all 50 animals. It will be enough to sample $n = 38$ animals. In the latter case $b = 5$, $\alpha = 0$ and $\beta = 0.954$.

A3.5 A CROSS-SECTIONAL OR A LONGITUDINAL POINT OF VIEW?

Before we proceed with the discussion of relatively small herds, we will first turn to another related and important issue. Do we want to judge a farm on the basis of the present visit alone or do we take the track record of that farm into account as well? In other words: do we take a cross-sectional or a longitudinal point of view? In the latter case we regard a sample of animals collected during a visit to the farm as part of a long sequence of animals produced by the farm: a reflection of welfare of animals in an ongoing management process. The underlying population of animals consists of all animals that are and will be produced at that particular farm. The proportion p is the probability for any cow produced by the farm to be defective. With this longitudinal view point, results for infinite populations apply. If we take the cross-sectional point of view, and we base our conclusions on the current visit only, the population is the herd at the moment of the visit (50 animals in our example) and p is the proportion of defectives among these 50 animals. Results for finite populations apply. Note that the populations involved and the interpretation of proportion p of defectives are different under the two points of view. This difference between the two approaches is important, both with respect to our monitoring strategy and with respect to the sample sizes we can practically acquire.

A3.6 ACCOUNTING FOR FINITE HERD SIZE (SMALL TO MODERATELY SIZED FARMS)

For a small farm of say 50 animals, from the cross-sectional point of view we observed (at the end of Section 4) that we need to sample $n = 38$ animals. But, from the longitudinal point of view, according to the results in Section 2 (true proportion $p = 0.2$, top half of the table) we need to sample as many as $n = 135$ animals. At first sight there is something of a contradiction here. The difference between the sample sizes lies in the different populations assumed and the different interpretation of p. If we sample all 50 animals and we detect $x = 6$ defective animals, we know that at the moment of the visit the proportion of defectives exceeds 0.10. So, it seems that we are right in giving a warning to the farmer. However, if we take the longitudinal point of view, it is quite possible for the proportion p in the long run to be smaller than 0.10. There will be variation in the number of defectives present at any moment and the 6 defectives detected during our visit may just be there by chance. May be the management is in good order. May be at previous visits no defectives were detected at all. We were just visiting the farm at the wrong moment. The number of animals present at our current visit (e.g. 50) is not large enough to offer a reliable impression of the performance of the farm in the long run. Hence, from a longitudinal point of view, no warning should be given.

So, what should a warning reflect: (1) the welfare state of the farm in the long run (longitudinal point of view) or (2) the welfare state of the farm at the moment of the visit (cross-sectional point of view)? Probably a critical discussion with the farmer about the presence of $x = 6$ defectives would be in good order (and possibly an explanation can be found), but the loss of a 'Michelin star' should perhaps be based on a longitudinal point of view, taking previous visits into account as well.

A3.7 FREQUENCY OF VISITS

The longitudinal view point seems quite natural. It seems reasonable that we look more critical at farms with a poor track record, may be by increasing the frequency of visits or by planning second visits for confirmation. Likewise, we might decide to visit farms with an excellent track record less frequently. So, if we are willing to postpone a definite warning (may be involving the loss of a 'Michelin' star) until several visits are performed, the longitudinal view point allows us to acquire larger (pooled) sample sizes and offer more reliable warnings.

Note that for the longitudinal point of view we can adopt more intricate procedures. For instance, we may combine results from several visits with larger statistical weight for more

recent visits, similar to the use of a moving average. We will not go into this any further, because it does not affect decisions we have to take at this stage of the project. The most important thing now to decide upon is whether emphasis should be placed on the longitudinal or on the cross-sectional point of view. From the longitudinal view point critical comments may be offered on the basis of one visit only, but 'loss of stars' should be based on several visits. In order to distribute our limited resources (number of observers, frequency of visits) over the farms, a longitudinal view point with more focus on farms with a poor track record could make sense.

A3.8 THE CROSS-SECTIONAL VIEW POINT FROM A PRAGMATIC PERSPECTIVE

Here we focus on a single visit and on the true proportion of defectives p that is present at that visit. Suppose that we determine, taking account of all practical constraints (such as time and money), how many animals we can inspect during a single visit. So, n is the number of animals that we can afford to inspect rather than the number of animals that is required. What kind of departures from the critical threshold are we able to detect with such a sample size?

For illustrative purposes we will stick to the critical threshold of 0.10. The herd size is $N = 150$. The sample size is $n = 50$. In the table below, a number of values for the true proportion p are listed that all exceed the critical threshold of 0.10. For each value, the probability β of detecting the departure from 0.10, i.e. the probability of rejecting the hypothesis $p \leq 0.10$, is calculated. So, the power β of the statistical test for various alternatives of the hypothesis $p \leq 0.10$ is calculated. In all instances the size α does not exceed 0.05. Because the observations x are counts, it is possible for α to be markedly below 0.05.

Table A3.2 shows what kind of departures from the critical threshold we are likely to detect. For instance, when $p = 0.17$, which means that there are 25 defectives on the farm, we have a probability of 0.46 to detect the difference with the threshold of 0.10.

Some results for a larger herd size $N = 300$ are presented in Table A3.3.

Possibly because in the last case a size α that is much closer to 0.05 can be achieved, the results for power β for the two herd sizes are not that different.

TABLE A3.2 Calculation of the power β = probability to detect that the true proportion of defectives is larger than .10 (N=150, n=50).

True p	.13	.17	.20	.25	.30
Number of defectives in the population	20	25	30	37	45
Power β	.17	.46	.74	.94	.99

Notes: herd size is taken into account in the calculations; the probability β depends on the true proportion p of defectives in the population; it also depends on the herd size (N) and the sample size (n); we decide that the proportion of defectives exceeds 0.1 when the number of defectives in the sample (x) exceeds value b = 8; value b = 8 follows from N, n, critical threshold .10 and the requirement $\alpha < .05$; the size of the (one-sided) test is α = .024.

Table A3.3 Calculation of the power β = probability to detect that the true proportion of defectives is larger than .10 (N=300, n=50).

True p	.13	.17	.20	.25	.30
Number of defectives in the population	39	51	60	75	90
Power β	.18	.49	.71	.93	.99

Notes: herd size is taken into account in the calculations; we decide that the proportion of defectives exceeds .1 when the number of defectives in the sample (x) exceeds value b = 8.; again, value b = 8 follows from N, n, critical threshold 0.10 and the requirement $\alpha < 0.05$; the size of the (one-sided) test is α = .042.

A3.9 WHAT IS NEEDED FOR A POWER CALCULATION?

In order to be able to calculate the power β in other instances, we need:

1. The critical threshold (0.10 in this note).
2. The herd size N (150 and 300 in the examples).
3. The sample size n (50 in the example).
4. The departure from the critical threshold (e.g. p = 0.20).

If a range of herd sizes is contemplated, a reasonable lower and upper bound and a mean or median value are required. Likewise calculations can be performed for several alternative values for the critical threshold (within reason).

A3.10 ACCURACY OF PROPORTIONS

This note primarily focuses on accuracy required for estimated proportions such that (1) the probability for a false warning is small (below 0.05) and (2) the probability to detect a departure from a critical threshold is large (above 0.95). In addition, on the basis of estimated proportions, we would also like to assess the variation between herds and possibly compare differences between some of the herds. To that end we have a closer look at the accuracy of an estimated proportion of defectives in a herd.

For any observed proportion x/n we can construct 95% confidence limits for the associated true proportion. For example, if we adopt the longitudinal point of view (infinite population size) and the sample size $n = 135$ (based on several subsequent visits say), while $x = 108$, so $x/n = 0.8$, these 95% confidence limits are: 0.72 and 0.86. The width of the interval indicates how accurate our statements about the true proportion can be. Roughly speaking, the true proportion would be between 0.72 and 0.86. This gives an idea of the kind of variation between farms, or differences between farms, we are able to detect. The interval will be wider, when the proportion is closer to 0.5. When the sample size is not too small, the confidence bounds may be roughly approximated by the following expressions:

$$\text{upper bound} = \frac{x}{n} + 1.96 * \sqrt{\frac{1}{n}\frac{x}{n}\left(1 - \frac{x}{n}\right)}$$

$$\text{lower bound} = \frac{x}{n} - 1.96 * \sqrt{\frac{1}{n}\frac{x}{n}\left(1 - \frac{x}{n}\right)}$$

For $x = 108$ and $n = 135$, the approximate bounds are 0.73 and 0.87, which is indeed close to the aforementioned exact results of 0.72 and 0.86. The approximate bounds are symmetric around the estimated proportion of 0.8. When the proportion is more extreme and / or the sample size is smaller, the exact interval will tend to be more and more asymmetric around the estimated proportion. For instance, for a small proportion, the interval will tend to be shorter on the left and longer on the right hand side of the estimated proportion. So, in that case the approximate interval (that is always symmetric) will differ more from the exact interval. The estimated standard error of the proportion is:

$$\sqrt{\frac{1}{n}\frac{x}{n}\left(1 - \frac{x}{n}\right)}$$

Thus, the approximate interval roughly equals the estimated proportion plus or minus twice the estimated standard error of the estimated proportion. Suppose that for another herd, $x = 35$, $n = 135$, so $x/n = 0.26$. For that herd, the exact bounds are 0.19 and 0.34, while the approximate bounds are 0.19 and 0.33. The fact that the intervals for the two herds do not overlap at all suggests that the proportions of defectives at the two farms are significantly different. An exact comparison may be based on Fisher's exact test, applied to the Table A3.4.

TABLE A3.4 Fisher's exact test.

	Defective	Non-defective
Herd 1	108	27
Herd 2	35	100

The numbers in Table A3.4 are corresponding to the numbers of defective and non-defective animals in the samples collected at the two farms. The P-value of the test is very small ($P < 0.001$), indicating that indeed the two proportions for the two farms are significantly different.

A3.11 FINAL REMARKS

In this note we have focussed on warnings based on apparent proportions. A similar discussion could be held about 'compliments' for excellent welfare, may be resulting in an extra 'Michelin star'. We have assumed that our judgement of animals is faultless. This assumption directly relates to the repeatability and reproducibility of observers (variation within an observer and variation between observers). Introduction of a sensitivity (probability that an observer recognises a truly defective animal) and specificity (probability that an observer recognises a truly sound animal) complicates matters considerably. For instance, when there is a possibility of a false positive (one minus the specificity), with an increasing sample size, more false positives will be introduced. With poor specificity it may simply not be possible, whatever the sample size n, to achieve $\alpha \leq 0.05$ and $\beta \geq 0.95$. We have also assumed that defectives are spread in a random fashion over the population. When there is clustering in time, a statistical evaluation is considerably more complicated. I have assumed that in case of clustering, there will usually be an acceptable explanation, such as a regional outbreak of a disease that makes further evaluation unnecessary. Of course data corresponding to such periods of calamity that are beyond control of the farmer should be no part of a longitudinal evaluation.

Various proportions will be part of the total monitoring system. In WP2.3 these proportions (and other measurements) will be combined into a number of summary scores associated with different broad aspects of welfare, such as animal health for instance. When a farm receives a poor score, we would like to identify the basic measurements that explain this poor score. For a proper advice to the farmer, we need adequate accuracy of estimated proportions. Adequate accuracy of estimated proportions seems a good starting point for good quality of the summary scores as constructed in WP 2.3 as well. However, it is very difficult, if not impossible, to obtain analytic results for accuracy of estimation of these summary scores. In principle, there is the possibility to run the monitoring system on simulated data, where the input for the simulations might be based on the large practical data set that will be collected in WP2.4. But this is not an easy exercise. It is possible to take half of the intended number of farms and double the number of visits, in order to

evaluate the repeatability of the summary measures from WP2.3 from the large practical data set .However, this would mean that effectively far less data would be available for our primary goals in WP2.4, namely quantifying associations between measurements on animals or pens, associations between measurements on animals or pens and farm characteristics, finding risk factors and slimming down the full monitoring system. It would not be unreasonable to state that a direct validation of the summary scores of WP2.3 on the basis of the practical data set collected in WP2.4 is beyond the present resources of the project, but should be part of a future evaluation.