Implementation of the Welfare Quality® broiler assessment protocol – final report

Overview of outcome-based measurement of broiler welfare and a general discussion on the Welfare Quality® broiler assessment protocol

Ingrid C. de Jong, Henk Gunnink and Vincent Hindle

LIVESTOCK RESEARCH
WAGENINGEN UR

Implementation of the Welfare Quality® broiler assessment protocol – final report

Overview of outcome-based measurement of broiler welfare and a general discussion on the Welfare Quality® broiler assessment protocol

Ingrid C. de Jong, Henk Gunnink and Vincent Hindle

This research was conducted by Wageningen UR Livestock Research, commissioned and funded by the Ministry of Economic Affairs, within the framework of Policy Support Research theme
'Animal Welfare' (project number BO-20-008-002.05)

Wageningen UR Livestock Research
Wageningen, January 2014

Livestock Research Report 833

LIVESTOCK RESEARCH
WAGENINGEN UR

De Jong, I.C., 2014. *Implementation of the Welfare Quality® broiler assessment protocol – final report. Overview of outcome-based measurement of broiler welfare and a general discussion on the Welfare Quality® broiler assessment protocol* Wageningen, Wageningen UR (University & Research centre) Livestock Research, Livestock Research Report 833. 58 blz.

Samenvatting Dit rapport bediscussieert twee onderwerpen in relatie tot het onderzoek naar de implementatie van de Welfare Quality® vleeskuikenassessment protocol. Als eerste worden dierkenmerken, die op dit moment onderdeel zijn van het Welfare Quality® vleeskuikenprotocol, besproken met betrekking tot de geschiktheid om als doelkenmerk te worden opgenomen in toekomstige wet- of regelgeving. De beoordeling van de dierkenmerken is gedaan op basis van ervaringen met het meten van de diverse kenmerken en een literatuurstudie. Vervolgens wordt de welzijnsassessment protocol besproken in relatie tot een vereenvoudiging, de (on)bruikbaarheid van de indicatoren en het integratiemodel, en wordt een advies gegeven met betrekking tot de toepassing van de assessment protocol in de praktijk.

Summary In the current report, two topics related to our study on the implementation of the Welfare Quality® broiler assessment protocol are discussed. First, animal-based measures for broiler welfare, currently included in the Welfare Quality® broiler assessment protocol, are discussed with respect to their suitability as outcome-based measure in future legislation or regulations. Practical experiences with the measures and supporting literature are used to judge the various animal-based measures. Second, the Welfare Quality® broiler assessment protocol is discussed with respect to a simplification, the relevance of the various measures and the aggregation model, and it is suggested how the assessment protocol can be applied in practice.

Livestock Research Report

Contents

Preface

The following people are acknowledged for their contribution to the project: Vincent Hindle, Henk Gunnink, Hans van den Heuvel, Piet van Wikselaar, Ruud Dekker, Henk Schilder, Gisabeth Binnendijk, Theo van Hattum, Ido Alferink and Lizzy Siersma. The project was financed by the Ministry of Economic Affairs and partially subsidised by the Product Board of Meat and Eggs. We thank all farmers that allowed us to perform the assessments on their flocks, the slaughter plants for the cooperation in data collection and the veterinarians (farm advisors) for their contribution to the project.

Ingrid de Jong (project leader)

Summary

The current report presents the final results of the project 'Implementation of the Welfare Quality® (WQ) broiler assessment protocol'.

First, the potential for use of animal-based measures of the WQ broiler assessment protocol are discussed in relation to their suitability as outcome-based measure in future legislation. In addition to measures indicative of on-farm welfare, we also discussed measures indicative of broiler welfare during transport, stunning and killing. Our experience with the measures in the WQ broiler assessment protocol and additional literature study were used to evaluate the measures. Animal-based measures were evaluated with respect to three criteria: (1) the relationship with broiler welfare; (2) risk factors for the specific welfare issue, and thus if there are tools for improvement, and (3) feasibility in practice. In addition, we included a description of possible iceberg indicators for broiler welfare (iceberg indicator being defined as an indicator which provides an overall assessment of welfare and effectively summarises several measures of welfare). Animal-based measures for broiler welfare that can (possibly) be used as outcome-based measure in future legislation are footpad dermatitis and hock burn, mortality (especially if culling, first week mortality and total mortality are registered separately), dead-on-arrival and fractures and bruises. Potential iceberg indicators are the percentage of rejections, dead-on-arrival, and breast irritation/blisters. Few important welfare indicators are not feasible in practice when these should be included in routine measures for legislation, such as gait score to assess lameness, which is a very time consuming procedure. With respect to behaviour, which is an important welfare aspect, there is a need to develop valid and feasible animal-based measures.

In the second chapter of the report we discussed the WQ broiler assessment protocol in relation to (1) possibilities for simplification (making it less time consuming), (2) the relevance of the measures in relation to broiler welfare, and (3) the aggregation model to calculate a final flock score and assign a flock to one of four welfare categories defined by Welfare Quality®.

Reduction of assessment time of the WQ broiler protocol can be achieved by using slaughter plant measures of footpad dermatitis and hock burn to predict the clinical scores (cleanliness, footpad dermatitis and hock burn) on farm. This leads to a reduction in assessment time by approximately one hour.

The following animal-based measures that are included in the WQ broiler assessment protocol have been shown to be feasible and valid: footpad dermatitis, hock burn, breast blister, lameness, mortality and culls, rejection figures, cleanliness, panting and huddling. It is advised to improve the measures of rejection and mortality by more specifically scoring rejections and separate scoring of culls. The following animal-based measures as currently described in the WQ protocol need to be improved or validated: breast blisters, Qualitative Behaviour Assessment, Touch test. With respect to several resource-based measures currently included in the WQ protocol, it is advised to develop animal-based measures where possible, especially with respect to criterion 10 (expression of other behaviour). In addition, because appropriate behaviour is an important welfare aspect it is strongly advised to give priority to the development of valid animal-based measures for this principle, as there is currently no valid animal-based measure available for broiler chicken behaviour.

Considerations that need to be taken into account before the WQ broiler assessment protocol can be applied in practice are (1) the visit frequency; and (2) providing feed-back based on individual measures instead of using the aggregation method as proposed by Welfare Quality®, because the aggregation method in its present form has been shown to be insensitive to actual differences in welfare level between flocks. In addition, for a few welfare measures management tools for farmers to be able to improve their scores are scarce (e.g., with respect to lameness which is largely related to

the genetic background of the birds). If (parts of) the WQ broiler assessment protocol is to be implemented in practice as a tool for management, farmers need to be facilitated with sufficient tools for improvement.

In conclusion, parts of the WQ broiler assessment protocol are suitable as management tool for farmers to improve the welfare of broiler chickens on-farm. It is advised to start with the feasible and valid animal-based measures. At the same time, animal-based measures should be developed where currently in the WQ broiler protocol information is lacking or invalid, in order to include more welfare aspects in the broiler assessment protocol in the future. In addition, if overall flock classification is required, the aggregation method will have to be improved so that actual differences in welfare level between flocks are reflected in the final flock classifications.

1 Contents of the report

This report presents the final results of the project 'Implementation of the Welfare Quality® (WQ) broiler assessment protocol' with respect to two topics:

- potential for use of the animal-based measures of the WQ broiler welfare assessment protocol as outcome-based measures for broiler welfare, based on WUR-LR experience with the Welfare Quality® broiler assessment protocol (chapter 2);
- a general discussion regarding the Welfare Quality® broiler assessment protocol as indicator of broiler welfare on-farm: possibilities for simplification, the relevance of current protocol measures and the Welfare Quality® aggregation model, including suggestions on how the assessment protocol can be applied in practice (chapter 3).

2 Outcome based measurement of broiler welfare

2.1 Introduction

Animal welfare legislation or regulations, commonly rely on management- or resource-based measures instead of outcome-based/animal-based measures. Thus, an animal welfare indicator requires information concerning house dimensions, stocking densities, light intensity or lighting regime programmes. Management- or resource-based measures are usually relatively easy to check by the inspection authority (e.g., in farm administration). However, animals differ in their genetic inheritance, early-life experience and temperament and consequently may experience their environment in different ways. Apparently similar environments may be managed differently by the stockperson, further affecting the animals' experience of a particular situation. Therefore, resource- or management-based measures may provide only partial information about the animals' welfare in particular situations. Outcome-based/animal-based measures reflect the actual welfare state of the animal in terms of their behaviour, health, fearfulness, physical condition etc. The fact that such outcome-based/animal-based measures are sensitive to farm management and specific system-animal interactions makes them particularly relevant indicators of animal welfare (Blokhuis et al., 2010).

The current Broiler Directive (2007/43/EC) is an exception with respect to the use of outcome-based measures. If broilers are housed at the maximum stocking density, a farmer should fulfil the mortality requirements with respect to the chickens. In the Netherlands, additional requirements for footpad dermatitis have been introduced ('Afsprakenkader Vleeskuikenrichtlijn'; http://edepot.wur.nl/12400). It is up to the farmer to decide how to manage and house his flock in line with the requirements with respect to mortality and footpad dermatitis.

Application of outcome-based measures in legislation or regulations has several advantages, i.e. allowing a farmer the freedom to determine his own management approach while providing a more reliable measure of animal welfare than with resource- or management-based measures (Blokhuis et al., 2010). However, use of outcome-based measures in legislation has disadvantages: inspection might be more difficult or time-consuming for measurement of outcome-based measures. Farmers might struggle with a lack of 'rules' on how to keep and manage their flock properly.

Wageningen UR Livestock research (WUR-LR) has recently performed the project 'Implementation of the broiler welfare assessment protocol'. Stakeholders involved in this project have required a discussion of the potential use of the measures of the broiler welfare assessment protocol as outcome-based indicator of broiler welfare, based on WUR-LR experience with broiler protocol (De Jong et al., 2011; De Jong et al., unpublished observations). The aim of this chapter is to provide an overview of the potentials for use of outcome-based measures, instead of resource or management based measures, based on experiences with the Welfare Quality® broiler protocol. In addition, relevant information from other research projects and scientific literature has been taken into account.

2.1.1 Definitions and terms

A list of terms and definitions used in the current report is provided below:

- Outcome-based measure: describes the goal that needs to be achieved (animal welfare) and relate it to the person concerned (farmer) how this goal can be achieved.
- Resource based measure: measure taken with regard to the environment in which the animals are kept (Welfare Quality, 2009). For example the number of drinkers or feeder space.
- Animal based measure: measure that can be taken directly from the animal (Welfare Quality, 2009). For example behaviour or wounds.

- Management-based measure: measures that refer to what the animal unit manager does on the animal unit and which management processes are used (Welfare Quality, 2009). Examples include feeding schedules or lighting programmes.
- Iceberg indicator or key indicator: An 'iceberg' indicator provides an overall assessment of welfare, just as the protruding tip of an iceberg indicates the submerged bulk beneath the water's surface. It effectively summarises several measures of welfare and is easy to understand (FAWC, 2009).

2.2 Methods

2.2.1 Screening of measures in the Welfare Quality® broiler assessment protocol

The Welfare Quality® broiler protocol (Welfare Quality, 2009) uses several animal-based measures to evaluate broiler welfare on-farm, during transportation and at slaughter. Based on our experience with the Welfare Quality® broiler assessment protocol, the various animal-based measures were evaluated with respect to their suitability to be used as outcome-based measure. Each measure was evaluated according to the following questions:

1. It's relationship with broiler welfare;
2. Risk factors for the specific welfare problem, and thus tool for improvement;
3. The currently used method(s) of measurement (feasibility);

Additionally, based on the answers to:

4. Concluding remarks: suitability as outcome-based measure for broiler welfare.

In addition to the experiences with the Welfare Quality® protocol, experiences gained within the 'Animal Welfare Check Points' project were used in the evaluation of various measures indicative of welfare during catching, transport and slaughter (Visser et al., 2014).

2.2.2 Literature study

A study of supporting literature concerning various animal-based measures and measures not included in the broiler protocol (but potentially useful as animal welfare indicator) was also performed.

With respect to some of the animal-based measures, many relevant studies are available (e.g. on footpad dermatitis). In the current report only the most relevant studies have been included.

2.2.3 Iceberg indicators

Iceberg or key indicators effectively summarise many measures of welfare and are relatively easy to understand. Iceberg indicators are outcome-based measures, but do have some specific characteristics. A description of possible iceberg indicators for broiler welfare was made for the Ministry of Economic Affairs and is included in Appendix 2.1 of the current report because of its relevance with respect to the discussion in the current report.

2.3 Results

2.3.1 Welfare Quality® broiler welfare assessment protocol

The Welfare Quality® consortium published the broiler welfare assessment protocol in 2009 (Welfare Quality, 2009). This welfare assessment protocol, like the assessment protocols that were developed for pigs, dairy cattle, veal calves and laying hens, focuses on animal-based measures for welfare instead of resource- or management-based measures (Blokhuis et al., 2010). Table 1 provides an overview of the measures that were included in the Welfare Quality® broiler assessment protocol to assess bird welfare on-farm. Subsequently, an assessment protocol was developed to assess the welfare of broiler chickens during transport and at slaughter (Table 2). Measures have been developed

to assess bird welfare during transport/slaughter, but as yet no method has been developed to aggregate the measures into a final flock score for welfare during transport and at slaughter.

Table 1

The Welfare Quality® broiler assessment protocol, to assess on-farm welfare (Welfare Quality, 2009)

Welfare principle	Welfare criterion	Measure[1 2]
Good feeding	Absence of prolonged hunger	*Emaciation (S)*
	Absence of prolonged thirst	Drinker space
Good housing	Comfort around resting	*Cleanliness,* litter quality, dust
	Thermaal comfort	*Panting, huddling*
	Ease of movement	Stocking density
Good health	Absence of injuries	*Lameness, hock burn (F+S), footpad dermatitis(F+S), breast blisters (S)*
	Absence of disease	*Mortality, culls on farm (S), pericarditis (S), septicaemia (S), hepatitis (S), dehydration (S), abcesses (S)*
	Absence of pain induced by management procedures	-[5]
Appropriate behaviour	Expression of social behaviours	-
	Expression of other behaviours	Cover on the range, free range
	Good human-animal relationship	*Touch test*[3]
	Positive emotional state	*Qualitative Behaviour Assessment*[4]

[1]Measures in italics are animal-based measures. Other measures are management- or resource-based measures.

[2]Measures indicated with (S) are measured during slaughter; measures indicated with (S+F) can be measured either on-farm or at slaughter.

[3] The touch test measures the number of birds within 1 m distance of the observer at various locations in the house (Welfare Quality, 2009).

[4] The Qualitative Behaviour Assessment (QBA) scores the behaviour of the flock using 23 descriptors (Welfare Quality, 2009)

[5] Empty cells indicate that there is yet no measure available for this criterion

Table 2

The Welfare Quality® broiler assessment protocol to assess welfare during transport and slaugther (Welfare Quality, 2009).

Welfare principle	Welfare criterion	Measure[1]
Good feeding	Absence of prolonged hunger	Feed withdrawal time
	Absence of prolonged thirst	Water withdrawal time
Good housing	Comfort around resting	-[2]
	Thermaal comfort	*Panting on lorry and/or lairage*
	Ease of movement	Stocking density in crates
Good health	Absence of injuries	*Wing damage, bruising*
	Absence of disease	*Dead-on-arrival*
	Absence of pain induced by management procedures	*Pre-stun shock, effectiveness of stunning*
Appropriate behaviour	Expression of social behaviours	-
	Expression of other behaviours	-
	Good human-animal relationship	-
	Positive emotional state	*Wing flapping on the line*

[1] Measures in italics are animal-based measures. Other measures are management- or resource based measures.

[2] Empty cells indicate that there is yet no measure available for this criterion

At present, most experience has been gained with the broiler welfare on-farm assessment protocol (De Jong et al., 2011; De Jong, unpublished results). This assessment protocol (Table 1) includes a number of animal-based measures with potential use as outcome-based measure. Below, all animal-based-measures are discussed with respect to their suitability as outcome-based indicator for broiler welfare.

At present we have less experience with the measures listed in Table 2, as indicators of broiler welfare during transport and slaughter. Some of these measures were included in the project 'Animal Welfare Check Points' (Visser et al., 2014) and these experiences have been taken into account in the discussion below.

The following paragraphs include a brief discussion of each indicator with respect to the following criteria:

1. Its relationship with broiler welfare;
2. Risk factors for the specific welfare problem;
3. How to measure (feasibility);
4. Concluding remarks on its suitability as outcome-based indicator of broiler welfare.

2.3.1.1 Rejections at slaughter

Relationship with broiler welfare
The Welfare Quality® broiler assessment protocol uses rejection figures to estimate absence of chronic hunger and absence of diseases: emaciation, ascites, dehydration, septicaemia, hepatitis, abscesses, pericarditis. Birds with these symptoms are usually rejected at slaughter because they are not suitable for human consumption. In addition, these measures are related to the prevalence of chronic hunger, diseases and trauma which are clearly a welfare concern in broiler flocks (EFSA, 2013a).

Risk factors
Risk factors for emaciation and dehydration are bird size (too small birds are unable to reach the feeders and drinkers, as height usually increases with increasing size of birds in a flock), lameness or presence of disease and trauma due to catching and transport procedures (EFSA, 2013). On-farm and flock information (combination of various factors including flock size, breed, stocking density, mortality, light intensity, water quality, ventilation type) explained 40% of the variation in rejection rate in a study in French broiler flocks (Lupo et al., 2010).

How to measure
In (almost) all countries the official authority (veterinarian) indicates the numbers of birds with the above mentioned reasons for rejections for each flock that is slaughtered (post mortem inspection). The method of registration however differs between countries: in The Netherlands the total number of rejected birds is registered, but not the number for each specific cause of rejection, in contrast to other countries. An example of the Dutch scoring form is shown in Appendix 1.

Suitability as outcome based indicator
Rejection rates (either in total or for specific reasons for rejection) are currently used as outcome-based indicator; if e.g. levels are above pre-defined trigger levels additional action will be undertaken. However, rejection rates are not very specific and may relate to various welfare issues, such as disease (on-farm) or trauma (related to the catching and transport process).

2.3.1.2 Plumage cleanliness

Relationship with bird welfare
Birds use their feathers to keep warm and to protect themselves from moisture and dirt and skin infections. Severe soiling with either dirt or faeces affects the protective properties of the feathers and thus has a negative effect on broiler welfare (Welfare Quality, 2009).

Risk factors
The most important risk factor to plumage cleanliness is a deterioration in litter quality. In addition, diarrhoea might negatively affect cleanliness. Bird activity and walking ability determine the time that the birds are in contact with the litter, and thus influence the risk for dirty feathers when litter quality deteriorates. It has been shown that there is a correlation between cleanliness and the prevalence of hock burn and footpad dermatitis in fast growing broiler strains (e.g., De Jong et al., 2011; De Jong, pers. comm.); plumage cleanliness, footpad dermatitis and hock burn are all related to the same risk factor (litter quality deterioration).

How to measure
Welfare Quality® assesses plumage cleanliness in a sample of at least 100 birds, taken from at least five different locations throughout the broiler house. It uses a four –class scoring system (clean – slightly dirty – moderately dirty – extremely dirty) by assessing the cleanliness of the feathers on the belly (Welfare Quality, 2009). In standard broilers, the belly is usually not well feathered, which might impair the assessment of cleanliness. In case of moderately to extremely dirty feathers the wing-, back- and leg feathers are usually also soiled. For white feathered birds, plumage cleanliness might be assessed at the plant using automated scoring systems; such a system has yet to be developed. Plumage cleanliness is currently not a standard measure at the slaughter plant.

Suitability as outcome-based indicator
Plumage cleanliness is mainly related to litter quality in the broiler house, which is clearly affected by the management of the farmer. Additionally, its negative effect on broiler welfare is not disputed. It might therefore be suitable as outcome based indicator. However, it should be noted that it has a common risk factor with contact dermatitis, and the additional value of including plumage cleanliness as outcome-based measure in addition to contact dermatitis might therefore be questioned. In addition, although it might be possible to assess plumage cleanliness at the plant and/or via automated assessment, these methods need to be developed further.

2.3.1.3 Panting and huddling

Relationship with broiler welfare
Panting and huddling are an indicator of the climate in the broiler house. Panting is defined as breathing rapidly in short gasps (Welfare Quality, 2009) and indicates that the environmental temperature is too high. Panting is a natural response of the bird to a too high environmental temperature, but persistent panting indicates that the climate in the house is not comfortable to the birds. Huddling is defined as a group of birds clumping together to preserve heat. It is different from clustering of birds because of synchrony in resting behaviour. Huddling can be observed in young chickens that are unable to maintain the body temperature by themselves (until 14 days of age) but is uncommon in birds near slaughter age. Both huddling and panting are stressors for a bird and have a negative effect on broiler welfare.

Risk factors
Environmental conditions i.e. room temperature , humidity and ventilation influence both panting and huddling. Even when enviromental conditions in a house are only slightly above normally acceptable levels, panting may be observed because the microclimate at broiler level is inadequate. In addition, a high stocking density, high growth rate and a high body weight can also influence panting behaviour (EFSA, 2010a). In a Dutch study, panting was observed on average 10% in flocks with fast growing strains and 2% in flocks with slow growing strains (De Jong et al., 2011). Experiences from another study indicate that the prevalence of panting is much higher when outside temperatures are above 25°C, despite the use of vapour (misting) systems in the houses (De Jong, pers. comm.). Huddling is very uncommon in indoor broiler housing but may occur under cold conditions (systems with outdoor ranges) or when there is a draught (De Jong, pers. comm.).

How to measure
Both panting and huddling are measured by estimating the number of birds showing this behaviour at five different locations throughout a broiler house (Welfare Quality, 2009). Both behaviours are easily and quickly (20 min per house) scored by a trained observer (De Jong, pers. comm.). It is recommended that it be investigated whether automated measurement of the distribution of broilers in a house are indicative of the climate at broiler level and whether or not scoring can be automated.

Suitability as outcome based indicator
Because huddling is not often observed it is questionable whether or not huddling is a suitable outcome-based measure. Panting, on the other hand, has a negative effect on broiler welfare and is clearly related to the (micro)climate of the house, which is a management aspect. Therefore, panting might provide a more suitable indicator for broiler welfare.

2.3.1.4 Lameness

Lameness is the inability of the bird to use one or both limbs in a normal manner. This can vary in severity from reduced ability or inability to bear weight, to total immobility (Welfare Quality, 2009). Lameness can be infectious or non-infectious (developmental or degenerative) and can involve tendons, joints, ligaments and bones (EFSA, 2010). Lameness provides a welfare indicator because lame birds may suffer from pain, might become unable or restricted in performance of natural behaviours such as pecking and scratching in the litter, dust bathing and running, and eventually be unable to reach food and water. There is however little evidence to link severity of lameness with disease and thus pain (Sandilands et al., 2011).

In broiler production systems with fast growing strains a high prevalence of lameness has been observed, which includes a range from slightly reduced ability to bear weight up to total immobility (gait scores 3, 4 and 5). Ranges between 30% (Knowles et al., 2008) and 50% (De Jong et al., 2011)) for gait scores 3-5 have been reported. Therefore, EFSA (2010) stated that lameness is one of the most important welfare problems in broiler chickens. Prevalence of lameness in slower growing strains is much lower (De Jong et al., 2011).

Risk factors

Body mass and growth rate can affect lameness, and for that reason there is a strong relationship with the genetic background of the broiler chicken (EFSA, 2010, Knowles et al., 2008). Other risk factors include lighting regime (shorter dark period during the day), a higher stocking density at the end of the rearing period, antibiotic use and the use of feed pellets instead of mash or crumbs (Knowles et al., 2008). However, genetic background is a much stronger risk factor than farm management (EFSA, 2010; Knowles et al., 2008).

How to measure

The most widely used assessment method for lameness in broiler chickens is the Bristol gait score (Kestin et al., 1992). Trained assessors assign a score between 0 (perfect gait) and 5 (unable to walk) based on a number of criteria. Although widely used, it has several disadvantages such as the need to be thoroughly trained, being time consuming (1.5 h for assessing 150 birds) and open to debate because of variation in interpretation of the different categories. However, no other methods have as yet been developed that can easily be applied in practice. Welfare Quality® assesses at least 150 birds at a minimum of six different locations throughout a broiler house
(Welfare Quality, 2009). At present a study is underway to investigate whether or not flock movement and distribution can be used to adequately determine lameness in broilers (Butterworth, pers. comm.).

Suitability as outcome based indicator

As lameness is considered to be one of the most important welfare problems in broilers, it might be preferred as outcome-based indicator of broiler welfare. However, there remains a difficulty concerning the feasibility of current measurement methods for lameness in practice (time consuming, substantial training necessary), and its strong relationship with genetic background and relatively low management control.

2.3.1.5 Hock burn

Relationship with broiler welfare

Hock burn is a dermatitis of the plantar surface of the hock of broiler chickens. Hock burn is visible as brown or black spots on the hocks of the birds and may range from mild to severe ulceration and skin inflammation (Greene et al., 1985). Hock burn is part of the syndrome of contact dermatitis, which also includes footpad dermatitis and breast irritation/blisters (Greene et al., 1985, Haslam et al., 2007, Martland, 1985) that are discussed below. Prevalence of footpad dermatitis is highest, followed by hock burn and breast irritation or blisters (Greene et al., 1985), De Jong et al., 2011). Severe forms of hock burn are considered to be painful for the bird and affect their ability to perform their natural behaviour.

On average, prevalence of severe hock burn in standard, fast growing broilers was found to be 10-15% and about 2% in slower growing strains (De Jong et al., 2011).

Risk factors
A common risk factor for all types of contact dermatitis (footpad dermatitis, hock burn and breast irritation or blisters) is poor litter quality (Hepworth et al., 2010), De Jong et al., 2014). Additional risk factors for hock burn are a high body weight and stocking density (kg/m^2) at two weeks of age; high stocking density and bird weight at slaughter age (5 weeks) (Hepworth et al., 2010). There is a relationship between gait score and hock burn score in standard, fast growing broilers. More hock burns are found in lame birds (De Jong et al., 2011); possibly caused by the fact that lame birds spend more time lying on their hocks, thereby increasing the risk of hock burn when litter quality has deteriorated.

How to measure
Hock burn is routinely assessed at the slaughter plant in several countries (Butterworth et al., in press) by manual scoring the hocks of the birds on the slaughter line. Although many scoring systems are available, Welfare Quality® uses a five-point scale to assess the severity of hock burn on-farm (live birds) and at the slaughter plant (Welfare Quality, 2009). In the Dutch situation, it was impossible to assess hock burn at the plant using a five-point scale due to the rapid line speeds. Therefore, it was decided to use a two-point scale, scoring either the absence (no discoloration or small (< 0.5 cm^2) brown or black spot) or presence of hock burn (black or brown spot > 0.5 cm^2) (De Jong et al., 2011). Although not currently applied, it might become possible to use video imaging systems to assess hock burn at the slaughter plant. Similar systems are currently in place for footpad dermatitis (De Jong, 2013) and the same methodology could be used to assess hock burn.

According to Welfare Quality® samples should be taken as at least 100 birds from five different locations throughout the broiler house, or two separate sample periods of five minutes per flock on the slaughter line which represents a sample size of at least 200 birds. The most severe score form both hocks is registered per bird sampled (Welfare Quality, 2009). Considering sample size, we refer to two EFSA reports on the methodology of data collection of welfare indicators in broiler slaughterhouses (EFSA, 2013a, 2013b).

Suitability as outcome-based indicator
Hock burn is currently measured at the slaughter plant in various countries (e.g., UK, Sweden, Denmark) by the government or slaughterhouse staff (Butterworth et al., in press). Hock burn is relatively easy to measure by a trained inspector at the slaughter plant and has a clear relationship with broiler welfare. It has a common risk factor with other types of contact dermatitis (deteriorated litter quality) but also some unique influencing factors.

2.3.1.6 Footpad dermatitis

Relationship with broiler welfare
Footpad dermatitis is a dermatitis of the plantar surface of the feet (Greene et al., 1985, Martland, 1985). In severe cases the erosions develop into ulceration with inflammatory reactions of the subcutaneous tissue. The lesions can become infected with a variety of bacteria and can even lead to joint inflammation (Berg, 1998). Such lesions can cause pain, whether infected or not, which constitutes a welfare issue.

Prevalence of severe footpad dermatitis in flocks of standard broiler chickens in The Netherlands in 2011 was 38.2% (de Jong et al., 2012a). Prevalence of severe footpad dermatitis in flocks with slower growing strains (but without outdoor range) was below 10% (De Jong et al., 2011). Reducing the prevalence of footpad dermatitis has been one of the main focus points of the Dutch broiler sector in recent years.

Risk factors
Deteriorated litter quality is a common risk factor for all types of contact dermatitis (EFSA, 2010b). Genotype, housing, age of birds (EFSA, 2010) and season (de Jong et al., 2012a) are additional risk factors.

How to measure
Foot pad lesions are currently included in broiler welfare legislation in various countries (e.g., The Netherlands, Sweden, Denmark) and the Broiler Directive (2007/43/EC) indicates their importance as welfare indicator. They can be scored in live birds on-farm but it is preferred to score these at the slaughter plant (de Jong et al., 2012b). Usually the lesions are scored in three classes (no lesions, mild lesions, severe lesions (Berg, 1998)) but Welfare Quality® applies the Bristol Foot Burn Scale assigning footpad scoring into five classes (Welfare Quality, 2009). A system for automated measurement on the slaughter line has been developed (De Jong, 2013) and is currently in place at some slaughter plants in The Netherlands, Denmark and Norway (De Jong, pers. comm.).

According to Welfare Quality® sample size should be at least 100 birds at five different locations throughout the broiler house, or two separate samples of 50 left or right feet each at the plant. For each live bird, the most severe score of both feet is registered (Welfare Quality, 2009); when sampling at the plant usually only the right or left foot is scored (De Jong, pers. comm.). Considering sample size, we refer to two EFSA reports on the methodology of data collection of welfare indicators in broiler slaughterhouses (EFSA, 2013a, 2013b).

Suitability as outcome-based indicator
Footpad lesions are currently included in broiler welfare legislation in various countries (e.g., The Netherlands, Sweden, Denmark) but not in all EU countries. Footpad lesions can be relatively easily assessed by a trained inspector or by an automated system at the slaughter plant and have a clear relationship with broiler welfare.

2.3.1.7 Breast blisters

Risk factors
As with the other types of contact dermatitis, the most important risk factor for breast irritation or buttons and breast blisters is poor litter quality (EFSA, 2010b). Additional risk factors have been found, such as stocking density and genotype (Allain et al., 2009), sex (Gouveia et al., 2009), feed composition (Rodenburg et al., 2008) and perch design (Nielsen, 2004).

How to measure
Breast irritation/buttons and blisters are preferably measured on the slaughter line. Usually, slaughter house staff or veterinary inspectors measure breast irritation and blisters as quality indicator of the chickens and this provides a reason for downgrading of carcasses or rejection (Butterworth et al., in press). According to the Welfare Quality® broiler protocol breast blisters and buttons are scored by observing the birds on the line during a period of 5 to 10 minutes (Welfare Quality, 2009). There might be possibilities for automated registration using video imaging systems, but such systems need to be developed.

Suitability as outcome-based indicator
Breast irritations/buttons and breast blisters have a common risk factor with hock burn and footpad dermatitis (poor litter quality) but there are some additional risk factors. As footpad dermatitis and hock burn develop at an earlier stage, these are more sensitive to deterioration in litter quality than breast irritation or blisters. Breast blisters may function as an iceberg-indicator, as these are only observed when litter quality is very bad (de Jong et al., 2014) or housing conditions are inadequate (such as perch design (Nielsen, 2004)). Breast irritation or blisters are currently measured at the slaughter plant in many countries, either by the government inspection authority or slaughterhouse staff, as indicator of broiler welfare and/or meat quality (Butterworth et al., in press).

2.3.1.8 Mortality and culls on-farm

Relationship with broiler welfare
The culling rate is registered by the farmer and is transmitted to the slaughterhouse with the flock according to the Broiler Directive (2007/43/EC) (EFSA, 2013b). Culling is the act to intentionally euthanatizing chickens for two main reasons (EFSA, 2010): i) if birds are unsuitable to be farmed (e.g.: wrong body conformation, surplus to requirement) - 'voluntary culling'; ii) or if, in production animals, animal welfare is seriously compromised (e.g. disease, sickness, injuries, lameness)- 'involuntary culling'. Involuntary culling is relevant to animal welfare. Mortality itself does not directly reflect animal welfare but can impact welfare if we consider the way and the reason an animal has died (e.g. diseases or lameness) (EFSA, 2010). In broilers only involuntary culling is relevant, voluntary culling seldom occurs.

Average mortality per flock was about 3% for standard and slower growing broilers (data collected in 2011) (De Jong et al., 2011).

Risk factors
Risk factors for mortality and culling include disease, injuries and lameness (EFSA, 2010). Chick quality is an important risk factor for first week mortality (EFSA, 2010).

How to measure
Both culling and cumulative daily mortality are registered by the farmer according to the Broiler Directive (2007/43/EC) (EFSA, 2013b). In flocks housed at stocking densities lower than 33 kg/m^2 the separate registration of culling is not obliged by legislation and therefore usually included in figures of total mortality (De Jong et al., 2011). Farmers often record first week mortality separately.

Suitability as outcome-based indicator
Mortality is currently included as outcome-based measure in the Broiler Directive. Mortality and culling are easy and simple to register for farmers. Culling and mortality should be as low as possible; when animals are sick or injured, culling is the best way to limit suffering. Even if the aim is to assessment protocol and reduce overall mortality, the number of animals found dead should be considered separately from the number of animals culled, since culling of sick birds is desirable to limit suffering (EFSA, 2013b). In this case, the ratio culled birds/birds found dead can be a useful indicator and can reflect good management (EFSA, 2010). On the contrary, an excessively high culling rate could indicate housing or management errors resulting in poor conditions for birds.

2.3.1.9 Behaviour

Relationship with broiler welfare
In the Welfare Quality® broiler assessment protocol animal-based measures are included for two aspects of behaviour: absence of fear for humans (human-animal relationship) and positive emotional state. Fear for humans may negatively affect productivity and causes stress which negatively affects broiler welfare (Hemsworth, 2003). Welfare is often measured as absence of any negative state (such as absence of fear), but also the positive emotional state ('pleasure') of an animal is important with respect to welfare. In the Welfare Quality® broiler assessment protocol this is assessed by using the Qualitative Behaviour Assessment, providing a general impression of how broilers within a flock interact with each other and their environment (Welfare Quality, 2009).

Risk factors
Rough handling of birds, a low frequency of human-animal contact, but also other factors such as light intensity, restricted water supply, absence of separated resting areas may affect the response of animals towards humans (De Jong et al., 2012c).

With respect to the positive emotional state, many factors may affect the behaviour of a flock. These are management and environmental factors such as stocking density, light intensity, feeding or water supply schedules, temperature, lay-out of the broiler house, stocking density, litter quality etc. (De Jong et al., 2012c).

How to measure
Although many methods are available to assess the behaviour of broiler chickens, the Welfare Quality® broiler assessment protocol thus far only includes two aspects of behaviour: the Touch Test to assess fear of humans and the Qualitative Behaviour Assessment (QBA) to assess the positive emotional state. There are currently no animal-based measures of social behaviour and expression of other behaviours included in the Welfare Quality® protocol, probably because methods to assess the behaviour are usually very time consuming and therefore less acceptable for inclusion in a assessment protocoling system.

There is currently debate about the validity of the Touch Test (De Jong et al., 2011; De Jong, pers. comm.) as well as the QBA (De Jong, pers. comm.). The measures for the principle 'appropriate behaviour' in the Welfare Quality® broiler assessment protocol require further attention.

There are automated systems that are able to register the activity of a flock during the rearing period. It should be further studied whether or not these can be used for automatic registration of behavioural aspects in a flock.

Suitability as outcome-based indicator
Because of the lack of knowledge on the validity of the Touch Test and QBA, and the debate on the validity of both tests (De Jong, pers. comm.), these measures are yet not suitable as outcome-based indicators of broiler welfare. Other measures of behaviour are usually very time consuming and therefore not feasible to be included in a assessment protocoling protocol, until alternative methods (such as automated measures) become available.

2.3.1.10 Panting on the lorry/in lairage

Relationship with broiler welfare
As with panting on-farm, panting in the lorry is an indicator of the micro-climate in the lorry during transport and in lairage. Panting is defined as breathing rapidly in short gasps (Welfare Quality, 2009) and indicates that the environmental temperature is too high. Panting is a natural response of the bird to a too high environmental temperature.

Risk factors
Risk factors are the stocking density in crates and the environmental temperature in the lorry. As thermal stress is related to growth rate which is in turn related to genetics, this may also be considered as a risk factor (EFSA, 2010).

How to measure
The Welfare Quality® protocol measures the number of birds showing panting in 20 crates distributed throughout the lorry, when the lorry is in lairage. In the WLR project "Animal Welfare Check Points", panting was observed after loading when the lorry was still on the farm. Only 3-4 crates could be scored (Visser et al., 2014). It is difficult to have a good view of all crates. Panting is easy to recognise for a trained observer.

Suitability as outcome-based indicator
A comfortable climate on the lorry, also when in lairage, is an important welfare aspect. Although manual registration is required, it is not very time consuming and easy to score.

2.3.1.11 Wing fractures and bruises

Relationship with broiler welfare
Fractures and bruising caused by catching, transport or unloading are painful for the bird and are indicative of poor handling and transport conditions. Recently, a prevalence of 3% for wing fractures was reported in a Dutch study. Prevalence of bruising was much lower (NVWA, 2014).

Risk factors
Catching and placement in the transport crates are critical activities to birds leading potentially to fractures and bruising. When caught by hand, usually several birds are held by the legs in one hand increasing the risk of injury, alongside how they are caught, and how they are placed in the transport crate. Mechanical catching can also prove risk full if not performed correctly (Kettlewell and Mitchell, 1994). Unloading is also a potential risk full activity a risk resulting frequently in bruises and fractures dependent on the method of unloading (unloading by hand or tilting of the crates)(Kettlewell and Mitchell, 1994).

How to measure
Fractures and bruises can be scored on the slaughter line after plucking. Usually birds are observed in one or two samples per flock during several minutes, so that a sample of at least 200 birds is scored (Welfare Quality, 2009). Slaughter house personnel or government veterinarians measure the frequency of bruises and fractures as part of the total rejection, leading to carcass downgrading and is indicative of poor welfare during transport and catching on-farm (Butterworth et al., in press).

Suitability as outcome-based indicator
Although fractures and bruises are included in the rejection rate, specific figures are not calculated (VWA, 2014). Because bruises and fractures are indicative of welfare problems and can be scored on the slaughter-line, they are suitable as outcome-based indicators of broiler welfare during the catching and transport process. A trained observer, is able to distinguish between 'fresh' and 'old' wounds and thus to relate the prevalence to a specific stage of the broiler production chain (Reimert, pers. comm.), although it is also said that it is difficult to discriminate between fractures/bruising due to catching and transport or due to stunning, especially in the case of gas stunning.

2.3.1.12 Dead-on-Arrival (DoA)

Relationship with broiler welfare
Mortality during transport or in lairage is likely to be related to suffering; it is the ultimate outcome of broilers subjected to stressors (Lund et al., 2013) and therefore an indicator of welfare.

Risk factors for Dead-on-Arrival
Recent studies have shown that birds dead-on-arrival resulted predominantly from pre-slaughter handling (catching) and climatic conditions during transport (Chauvin et al., 2011, Lund et al., 2013). In addition, broilers not fit for transport have an increased risk of dying during transport (Nijdam et al., 2006); increased DoA has been associated with higher mortality levels on-farm (Chauvin et al., 2011). Although EFSA (2013) states that in slaughter-plants where gas stunning is used, the level of dead on arrival recorded could be underestimated as broilers are transferred directly in the crates from the truck to gas stunning system, impeding thorough observation of the birds upon arrival, others indicate that slaughter personnel is well able to select dead-on-arrival birds when hanging birds in the shackles.

How to measure
The number of birds dead-on-arrival is a standard measure at the slaughter plant. Figures are collected by the veterinary inspector of the government, slaughterhouse staff or the meat hygiene inspector (Butterworth et al, in press) when the birds are unloaded or shackled.

Suitability as outcome-based indicator
Dead-on-arrival figures are currently used as outcome-based indicator. When levels are excessively high, further action will be undertaken (Butterworth et al., in press).

2.3.1.13 Pre-stun shocks and unconsciousness

Relationship with broiler welfare
Inadequate stunning leads to suffering, any animal killed without being prior stunning, will be subjected to unnecessary to pain and distress.

Risk factors
Pre-stun shocks occur when a bird receives a premature electrical shock from the stunning bath, because it's head or wing touches water or wet surfaces at the very start of the entrance of the stunner (Welfare Quality, 2009). Inadequate stunning results in killing if conscious birds (Welfare Quality, 2009).

How to measure
Pre-stun shocks can be observed at the entrance to the water-bath stunner by counting the number of birds displaying avoidance, flapping or vocalisation. In the Welfare Quality® protocol usually one or more samples of several minutes are taken (Welfare Quality, 2009). At the stunner exit birds can be examined for signs of consciousness, including open eyes, neck arched with head directed vertically, wings held close to the body, extended legs or constant rapid body tremors, no rhythmic breathing. Usually one or more samples of a few minutes are taken (Welfare Quality, 2009).

Suitability as outcome-based indicator
Adequate stunning is important with respect to animal welfare. A trained observer can measure signs of consciousness and pre-stun shocks at the plant.

2.3.1.14 Flapping on the line

Relationship with broiler welfare
When birds wing flap they are just hang on the line, it indicates that the process is stress full to them and further injuries like fractures or bruises, can follow which are painful to the bird (Welfare Quality, 2009).

Risk factors
A risk factor is an inadequately designed slaughter line.

How to measure
According to Welfare Quality® birds should be observed just after 'shackling' for a couple of minutes and the number of birds flapping should be scored (Welfare Quality, 2009).

Suitability as outcome-based indicator
A properly designed plant is important with respect to welfare. A trained observer is able to measure wing flapping at the plant.

2.4 Other outcome-based measures for broiler welfare

There are a few other animal-based measures that are currently not included in the Welfare Quality® broiler assessment protocol but might be considered as outcome-based measure for broiler welfare:

- Feather damage and pecking injuries. Feather damage and wounds are included in the laying hen protocol, but not in the broiler protocol (Welfare Quality, 2009), probably because feather pecking is uncommon in broiler flocks. However, it can sometimes be observed in flocks with slower growing strains (De Jong, pers. comm.). Feather pecking and injuries have a negative effect on animal welfare.
- Thigh scratches. Scratches are painful and may be a route of infection. Thigh scratches are sometimes measured at the plant as indicator of meat quality (Butterworth et al., in press). Risk factors include, stocking density and bird activity, feeder and drinker space, light intensity and schedule (De Jong, pers. comm.). Automated measurement at the slaughter line using video equipment has not yet been developed but might be considered.
- Flock uniformity. Flock uniformity is considered to be an indicator of the conditions during the rearing period. Weights can be measured automatically on the slaughter line. Further research is necessary to determine whether flock uniformity is a suitable outcome-based welfare measure; it could be considered as an iceberg indicator.

2.5 Iceberg indicators

A definition of 'iceberg indicator' has been provided by FAWC (2009) and this definition has also been used by EFSA (2012): "An 'iceberg' indicator provides an overall assessment of welfare, just as the protruding tip of an iceberg signals its submerged bulk beneath the water's surface. It effectively summarises many measures of welfare and is easy to understand." Iceberg indicators are outcome-based and, like any measure of welfare, require scientific validation under farm conditions for all species. 'Iceberg indicator' is not a generally applied definition in scientific literature; some use the term 'key indicator' instead of 'iceberg indicator' (e.g., Kelly et al., 2011; 2013).

Advantages of using iceberg indicators to identify flocks or farms where welfare is at risk, are that measurement is usually simple and relatively cheap, and thus more feasible as compared to a (complete) welfare assessment protocol. In addition, iceberg indicators enable the identification of long-term trends in animal welfare. As for all welfare measures, iceberg indicators should be valid and repeatable and the interpretation should be unambiguous. FAWC (2009) indicates that especially productivity, total mortality, fertility and the use of mutilations are ambiguous. They suggest poor welfare but may also be related to an improvement of welfare in other areas (e.g. total mortality may indicate suffering through disease and dying, but may also be an indicator of a strict culling policy to avoid unnecessary pain and distress). Then their suitability as welfare indicator becomes questionable. In the ideal situation, iceberg indicators can be derived from data that are collected on a routine basis. For dairy cattle several studies have addressed the use of routinely collected data as iceberg indicator(s) (see e.g., Sandgren et al., 2009; Nyman et al., 2011; Kelly et al., 2011. 2013;De Vries et al., 2014).

Appendix 2.2 summarises possible iceberg indicators for broiler chickens.

2.6 Summary

In the results section various animal-based measures were described, taking into consideration the different aspects that may guide a decision to include these as outcome-based measures for broiler welfare. A final decision will also include other aspects, such as costs of measurement and the need for additional animal-based indicators to be included in (future) legislation.

In Table 3 the results of the screening of the various animal-based measures included in the Welfare Quality® broiler protocol are summarised. This table can be used as input for further discussion on the use of outcome-based measures for broiler welfare. Further developments, especially with automated measures, will create possibilities for the inclusion of more outcome-based measures of broiler welfare in legislation or regulations.

Certain new measures are proposed in addition to the measures already included in the Welfare Quality® protocol. Although these measures are potentially interesting, e.g. because these can be measured on the slaughter line, additional validation of these measures will be necessary.

According to Table 3, animal-based measures for broiler welfare that can be considered as outcome-based measure are: footpad dermatitis, hock burn, mortality, dead-on-arrival, fractures and bruises. Footpad dermatitis and mortality are already included as outcome-based welfare indicator in broiler welfare legislation in The Netherlands (Vleeskuikenrichtlijn) for flocks aimed to be kept at the highest stocking density (42 kg/m^2 with respect to footpad dermatitis and more than 39 kg/m^2 with respect to mortality). Rejection figures and breast blister/irritation might be suitable as iceberg indicator.

Table 3.

Summary of screening of various animal-based measures with respect to aspects that contribute to their suitability as outcome-based measure of broiler welfare, ranging from -- to ++. [1] This indicates whether or not the specific measure is a sensitive one, i.e. if changes in welfare on-farm/during transport and slaughter are reflected in outcomes of the measure (-- not sensitive to ++ very sensitive)

[2]This indicates if the risk factors for this welfare problem are known (--not known to ++ very well known);

[3]Indicates if methods have been developed, if these are not too time consuming (and expensive) or complicated (--no methods to 0 methods not feasible to ++ very feasible in practice;

[4]Indicates the possibility for automated measurement (on-farm or at slaughter) (--not possible to 0 possible but not developed to ++ possible and currently used);

[5]Summarizes the actual suitability as outcome-based measure (--not suitable until ++ very suitable).

Measure	Sensitive?	Risk factors known?[1]	Feasibility[3]	Automated measurement?[4]	Outcome-based measure?[5]	Additional remark
Rejections	0	+	++	--	0/+	Possible iceberg indicator
Cleanliness	+	++	-	0	-	Common risk factor with contact dermatitis
Panting and Huddling on-farm	+	++	0	-	-	
Lameness	++	++	--	0	-	
Hock burn	+	++	++	+	++	
Footpad dermatitis	++	++	++	++	++	
Breast blister/irritation	-	++	++	0	0	Possible iceberg indicator
Mortality on-farm	0	++	++	--	++	More sensitive in case culling, first week and total mortality are registered separately
Behaviour (QBA and touch test)	-	++	--	-/0	--	
Panting in lorry/lairage	+	++	0	--	0	
Fractures and bruises	++	++	+	+	++	
Dead-on-Arrival	0	++	++	--	++	Possible iceberg indicator
Pre-stun shocks and unconsciousness	++	++	0	--	0	
Flapping on the line	++	++	0	--	0	

2.7 Literature

Allain, V., Mirabito, L., Arnould, C., Colas, M., Le Bouquin, S., Lupo, C. & 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**,** 407-417.

Berg, C. C. 1998. Foot-pad dermatitis in broilers and turkeys - prevalence, risk factors and prevention. *Acta Universitatis Agriculturae Suaciae, Veterinaria 36.* Swedish University of Agricultural Sciences, Uppsala.

Blokhuis, H. J., Veissier, I., Miele, M. & Jones, B. 2010. The Welfare Quality (R) project and beyond: Safeguarding farm animal well-being. *Acta Agriculturae Scandinavica Section a-Animal Science,* 60**,** 129-140.

Butterworth, A., I.C. de Jong, C. Keppler, L. Stadig and S.L. Lambton, in press. Facilitation of communication on technical measures amongst competent authorities in the implementation of the Broiler Directive (2007/43/EC). Animal.

Chauvin, C., Hillion, S., Balaine, L., Michel, V., Peraste, J., Petetin, I., Lupo, C. & Le Bouquin, S. 2011. Factors associated with mortality of broilers during transport to slaughterhouse. *Animal,* 5**,** 287-293.

De Jong, I.C., 2013. Evaluatie Meyn footpad inspectie systeem. Livestock Research rapport 713.

de Jong, I. C., Gunnink, H. & van Harn, J. 2014. Wet litter not only induces footpad dermatitis but also reduces overall welfare, technical performance, and carcass yield in broiler chickens. *Journal of Applied Poultry Research,* 23**,** 51-58.

De Jong, I. C., Perez Moya, T., Gunnink, H., Van den Heuvel, H., Hindle, V., Mul, M. & Van Reenen, C. G. 2011. Symplifying the Welfare Quality assessment protocol for broilers. Report 533. Wageningen UR Livestock Research, Lelystad.

de Jong, I. C., van Harn, J., Gunnink, H., Hindle, V. A. & Lourens, A. 2012a. Footpad dermatitis in Dutch broiler flocks: Prevalence and factors of influence. *Poultry Science,* 91**,** 1569-74.

de Jong, I. C., van Harn, J., Gunnink, H., Lourens, A. & van Riel, J. W. 2012b. Measuring foot-pad lesions in commercial broiler houses. Some aspects of methodology. *Animal Welfare,* 21**,** 325-330.

De Jong I, Berg C, Butterworth A and Estevez I 2012c. Report updating the EFSA opinion on the Welfare of Broilers and Broiler Breeders – Part A – The Welfare of Chickens Kept for Meat Production. European Food Safety Authority, Parma, Italy.

EFSA, 2010. Scientific opinion on the influence of genetic parameters on the welfare and the resistance to stress of commercial broilers. *EFSA Journal,* 8**,** 1666.

EFSA, 2012. Statement on the use of animal-based measures to assess the welfare of animals. EFSA Journal 10 (6): 2767.

EFSA, 2013a. Sample Size Considerations for Hierarchical Populations. EFSA Journal 2013;11(7):3292-.

EFSA, 2013b. Technical assistance to the Commission (Article 31 of Regulation (EC) No 178/2002) for the preparation of a data collection system of welfare indicators in EU broilers' slaughterhouses. EFSA Journal 2013;11(7):3299-.

FAWC, 2009. Farm animal welfare in Great Britain: Past, Present and Future. 57 pp. Available from: http://webarchive.nationalarchives.gov.uk/20121007104210/http://www.fawc.org.uk/pdf/ppf-report091012.pdf

Gouveia, K. G., Vaz-Pires, P. & da Costa, P. M. 2009. Welfare assessment of broilers through examination of haematomas, foot-pad dermatitis, scratches and breast blisters at processing. *Animal Welfare,* 18**,** 43-48.

Greene, J. A., McCracken, R. M. & Evans, R. T. 1985. A CONTACT-DERMATITIS OF BROILERS - CLINICAL AND PATHOLOGICAL FINDINGS. *Avian Pathology,* 14**,** 23-38.

Haslam, S. M., Knowles, T. G., Brown, S. N., Wilkins, L. J., Kestin, S. C., Warriss, P. D. & Nicol, C. J. 2007. Factors affecting the prevalence of foot pad dermatitis, hock burn and breast burn in broiler chicken. *British Poultry Science,* 48**,** 264-275.

Hemsworth, P. H. 2003. Human-animal interactions in livestock production. Applied Animal Behaviour Science, 81, 185-198.

Hepworth, P. J., Nefedov, A. V., Muchnik, I. B. & Morgan, K. L. 2010. Early warning indicators for hock burn in broiler flocks. *Avian Pathology,* 39**,** 405-409.

Kelley, P.C., More, S.J., Blake, M., Hanlon, A., 2011. Identification of key performance indicators for on-farm animal welfare incidents: possible tools for early warning and prevention. Irish Veterinary Journal 64: 13.

Kelley, P.C., More, S.J., Blake, M., Higgins, I., Clegg, T., Hanlon, A., 2013. Validation of key indicators in cattle farms at high risk of animal welfare problems: a qualitative case-control study. Veterinary Record doi: 10.1136/vr.101177

Kestin, S. C., Knowles, T. G., Tinch, A. E. & Gregory, N. G. 1992. Prevalence of leg weakness in broiler chickens and its relationship with genotype. *The Veterinary Record,* 131**,** 190-194.

Kettlewell, P. J. & Mitchell, M. A. 1994. CATCHING, HANDLING AND LOADING OF POULTRY FOR ROAD TRANSPORTATION. *Worlds Poultry Science Journal,* 50**,** 54-56.

Knowles, T. G., Kestin, S. C., Haslam, S. M., Brown, S. N., Green, L. E., Butterworth, A., Pope, S. J., Pfeiffer, D. & Nicol, C. J. 2008. Leg Disorders in Broiler Chickens: Prevalence, Risk Factors and Prevention. *Plos One,* 3.

Lund, V. P., Kyvsgaard, N. C., Christensen, J. P. & Bisgaard, M. 2013. Pathological manifestations observed in dead-on-arrival broilers at a Danish abattoir. *British Poultry Science,* 54**,** 430-440.

Lupo, C., Bougeard, S., Balaine, L., Michel, V., Petetin, I., Colin, P., LeBouquin, S. & Chauvin, C. 2010. Risk factors for sanitary condemnation in broiler chickens and their relative impact: application of an original multiblock approach. *Epidemiology and Infection,* 138**,** 364-375.

Martland, M. F. 1985. ULCERATIVE DERMATITIS IN BROILER-CHICKENS - THE EFFECTS OF WET LITTER. *Avian Pathology,* 14**,** 353-364.

Nielsen, B. L. 2004. Breast blisters in groups of slow-growing broilers in relation to strain and the availability and use of perches. *British Poultry Science,* 45**,** 306-315.

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

NVWA, 2014. Rapportage project vervoer pluimveeketen 2013

Rodenburg, T. B., Van Harn, J., Van Krimpen, M. M., Ruis, M. A. W., Vermeij, I. & Spoolder, H. A. M. 2008. Comparison of three different diets for organic broilers: effects on performance and body condition. *British Poultry Science,* 49**,** 74-80.

Sandilands, V., Brocklehurst, S., Sparks, N., Baker, L., McGovern, R., Thorp, B. & Pearson, D. 2011. Assessing leg health in chickens using a force plate and gait scoring: how many birds is enough? *Veterinary Record,* 168**,** 77-77.

Visser, E.K., W. Ouweltjes, F. Neijenhuis, A. Lourens, J.T.N. Van der Werf, H. Gunnink, V.A. Hindle, J.C. Verkaik, G.P. Binnendijk en M.A. Gerritzen, 2014. Jaarrapportage onderzoek Animal Welfare Check Points 2013. Lelystad, Wageningen UR (University & Research centre) Livestock Research, Livestock Research Report 753, 90 blz.

Welfare Quality, 2009. The Welfare Quality assessment protocol for poultry (broilers, laying hens). The Welfare Quality Consortium, Lelystad.

Appendix 2.1. Rejection form broiler chickens

Bewijs van afkeuring

NR. [redacted]

Nederlandse Voedsel- en Warenautoriteit
Ministerie van Economische Zaken

[redacted] - vleeskuikens

Koppelidentificatie: [redacted]
Diersoort: vleeskuikens
Opzetdatum: [redacted]
Stal:

Van de aangeboden koppel is volgens opgave van het bedrijf:

Levend aangevoerd	25121	
Dood aangevoerd	15	(0,06%)
Geheel afgekeurd op ziekelijke afwijkingen	177	(0,70%)

De volgende afwijkingen zijn vastgesteld: +++ = Zeer veel; ++ = Veel; + = Weinig

Hele karkassen:

OIE ziektes	
Kleur / geur / afw. consistentie	+
Cachexie	
Hydrops	
(Poly-)serositis	+
(Poly-)arthritis / synovitis	
Hepatitis	+
Restverschijnselen O.r.	
(Ei)concrementen (kippen)	
Huidontsteking	
Open fracturen	
Omvangrijk letsel	
Overige	

Delen / organen:

Arthritis / synovitis	+
Restverschijnselen O.R.	
Huidontsteking	+
Open fracturen	
Omvangrijk letsel	+
Pericarditis	+
Salpingitis	
Afwijkende consistentie	
Overige	

Opmerking:

Datum: [redacted]
Plaats:

naamstempel toezichthouder

Handtekening:

Appendix 2.2. Table of possible iceberg indicators for broiler chicken welfare.

Table 2.1.

Promising indicators

Iceberg indicator	How to measure it?	Where / when to measure it?	What does it say about welfare?	How reliable is it?
Footpad dermatitis Included in Broiler Directive 2007/43/EC in some countries	Counting the number of feet with lesions and calculating the flock score[1] . See [6]	At the slaughter plant. Automated measures using video imaging systems is possible. Can also be measured on-farm, but that is less feasible and accurate[7]. Routinely measured in some countries (at the plant).	The number and severity of footpad dermatitis in a broiler flock is indicative of the general management on-farm: feed quality, water management, ventilation and house temperature, light regime, etc. and thus for bird welfare. In addition, FPD is painful for the bird and affects its ability to move (and eat and drink) [2,4], it is also a route for infection.	Relationship with management factors had been established in various studies, see e.g. [2,3,4,5]
Hock burn	Counting the number of feet with hock burn. See [6]	At the slaughter plant. Automated measures using video imaging systems is possible and has been developed for the UK (but impossible/needs to be adapted for e.g. NL because feet are removed by cutting the hocks). Can also be measured on-farm. Routinely measured in many countries as indicator of bird quality (at the plant).	Severe forms are considered to be painful and route for infection. The prevalence of hock burn is indicative of aspects of farm management, like footpad dermatitis[5], but there are more factors of influence such as bird weight and stocking density at two weeks of age. Prevalence of hock burn is related to fever and septicaemia, ascites, poorly grown birds, and abscesses, thus to general flock health [8]. In fast growing birds, there is a relationship between the quality of locomotion and the prevalence of hock burn [10 11]	Relationship with flock health and the other factors has been established in several studies, see e.g. [5, 8]
Cleanliness	See [6] for scoring method (four classes)	On-farm and (maybe) at the plant (before scalding). DK (?) routinely measures cleanliness at the plant.	Very dirty birds are indicative of bad litter quality, dirtiness is related to FPD and (to a lesser extent) hock burn [10, 11] and possibly also to flock activity. Dirty birds have problems with thermoregulation. Bad litter quality prevents the birds to perform its natural behaviour such as dustbathing and foraging.	Related to general farm management (temperature, ventilation, light regime, feed regime, etc). In general, related to the same risk factor as fpd and hock burn (bad litter quality)

Table 2.2.
Iceberg indicators that are less promising or are promising but need additional research with respect to validity/reliability/feasibility

Mortality Included in Broiler Directive 2007/43/EC	Counting the number of dead birds, including culling.	On-farm, daily registration by the farmer. If kept on the highest stocking density (> 39 kg/m^2) culling and mortality need to be registered separately (Broiler Directive); this is not commonly done for flocks kept at lower stocking densities.	Separation between culling and found dead is preferred. It is indicative of general flock health [9]. Interpretation might be difficult: flocks with e.g. locomotion problems where culling is applied might have better welfare but higher mortality as compared to flocks where no culling is applied.	There are various reasons for mortality and culling. Culling policy of a farmer largely affects the mortality rate.
Lameness	Applying the gait score to assess the locomotion of the broilers on-farm, see [6]. The gait score method is often discussed because of its reliability. It is not 1:1 related to the cause of lameness.	On-farm by measuring a sample of birds (e.g., 150 birds at least according to [6]). Usually measured just before sending the birds to slaughter. New methods for automated detection of flock activity are available and shown to be indicative of the gait score of the flock.	Lameness clearly affects welfare, either because it is painful, and/or because it affects the ability of the bird to perform its natural behaviour. Is related to the prevalence of hock burn in fast growing (but not slow growing) broilers [10]. Locomotion problems are amongst the most severe welfare problems for broiler chickens [9]. Is related to farm management, e.g. lighting program, feeding program, but largely affected by genetics.	Relationship with breed, day-old chicken quality and farm management. To a large degree related to bird genetics [9]
Rejections at slaughter (e.g., emaciation, abscesses, septicaemia, ascites etc) Included in the Broiler Directive 2007/43/EC	Routinely collected at slaughter by the veterinary inspection service. Usually an alert if the percentage exceeds a certain level.	At the plant. Usually done by the veterinary inspector.	General indicator of flock health. However, not very specific, methods/measures may differ between countries and a relation with other welfare measures has not been found [10,11]	Indicates a health problem if it exceeds a certain percentage. Measurements not very uniform between EU countries.
Thigh scratches	Sometimes routinely collected for assessing carcass quality. Assessment method: [4]	Preferably at the plant after plucking. Can also be measured on live birds on-farm.	Scratches are related to high stocking density and/or flock disturbance (which is related to housing design) [11]. Interpretation might be difficult, as inactive flocks due to leg problems might have little scratches but suffer from other welfare problems [4]	Needs further study

Wounds	See [6] (laying hen protocol)	At the plant, after plucking. Can also be measured on-farm.	Indicator of injurious pecking in a flock. This does not happen very often in broiler flocks, but sometimes occurs in slower growing flocks where birds are slaughtered at an older age. Injurious pecking is related to suboptimal management (see laying hen literature).	Reliable, see laying hen literature on injurious pecking
Flock activity	No standardised method available. Automated recording possible (see gait scoring). Or observations on-farm (time consuming)	On-farm.	The general activity level of a flock is related to the farm management, the breed and the quality of locomotion of the birds. It is not a very specific measure, but apathy, severe locomotion problems, disease, or panic reactions (fearful flocks) might be detected (if recorded automatically)	Needs more study
Flock uniformity	Usually measured at the plant as indicator of flock quality	Preferably at the plant. If automatic weighing, possibly on-farm.	Health problems, lameness, inadequate management, poor day old chick quality may lead to flocks with poor weight uniformity	Not used yet, but promising as possible welfare indicator. Needs further study.

1 Flock score FPD: FPD score = ((number of feet with score 1 * 0.5) + (number of feet with score 2 * 2))/Ntotal * 100
2 Shepherd, E. M. & Fairchild, B. D. 2010. Footpad dermatitis in poultry. Poultry Science, 89, 2043-2051.
3 de Jong, I. C., van Harn, J., Gunnink, H., Hindle, V. A. & Lourens, A. 2012. Footpad dermatitis in Dutch broiler flocks: Prevalence and factors of influence. Poultry Science, 91, 1569-74.
4 de Jong, I. C., Gunnink, H. & van Harn, J. 2014. Wet litter not only induces footpad dermatitis but also reduces overall welfare, technical performance, and carcass yield in broiler chickens. Journal of Applied Poultry Research, 23, 51-58.
5 Haslam, S. M., Knowles, T. G., Brown, S. N., Wilkins, L. J., Kestin, S. C., Warriss, P. D. & Nicol, C. J. 2007. Factors affecting the prevalence of foot pad dermatitis, hock burn and breast burn in broiler chicken. British Poultry Science, 48, 264-275.
6 Welfare Quality. 2009. The Welfare Quality assessment protocol for poultry (broilers, laying hens). The Welfare Quality Consortium, Lelystad.
7 de Jong, I. C., van Harn, J., Gunnink, H., Lourens, A. & van Riel, J. W. 2012. Measuring foot-pad lesions in commercial broiler houses. Some aspects of methodology. Animal Welfare, 21, 325-330.
8 Hepworth, P. J., Nefedov, A. V., Muchnik, I. B. & Morgan, K. L. 2011. Hock burn: an indicator of broiler flock health. Veterinary Record, 168.
9 EFSA. 2010. Scientific opinion on the influence of genetic parameters on the welfare and the resistance to stress of commercial broilers. EFSA Journal, 8, 1666.
10 De Jong, I. C., Perez Moya, T., Gunnink, H., Van den Heuvel, H., Hindle, V., Mul, M. & Van Reenen, C. G. 2011. Symplifying the Welfare Quality assessment protocol for broilers. Report 533. Wageningen UR Livestock Research, Lelystad.
11 De Jong, I.C., personal communication

3 The Welfare Quality® broiler assessment protocol: a general discussion

3.1 Introduction

In 2009 the Welfare Quality® consortium developed welfare assessment protocoling systems for various farm animal species, such as cattle, pigs, veal calves, laying hens and broiler chickens (Blokhuis et al., 2010). The Welfare Quality® (WQ) approach integrates the results of individual, preferably animal-based measures into an overall score for a farm or flock (see Figure 1). Individual measures are aggregated into scores for 12 welfare criteria, which are subsequently aggregated in scores for four welfare principles (Table 1). Finally a flock or unit score is assigned to one welfare classification category according to the principle scores the flock or unit attained (Welfare Quality, 2009). One application of such a assessment protocoling system is its integration in quality assurance schemes, for example. The WQ approach (from measures – to farm or flock score) might also be used as management support tool for a farmer, to stimulate improvement of animal welfare.

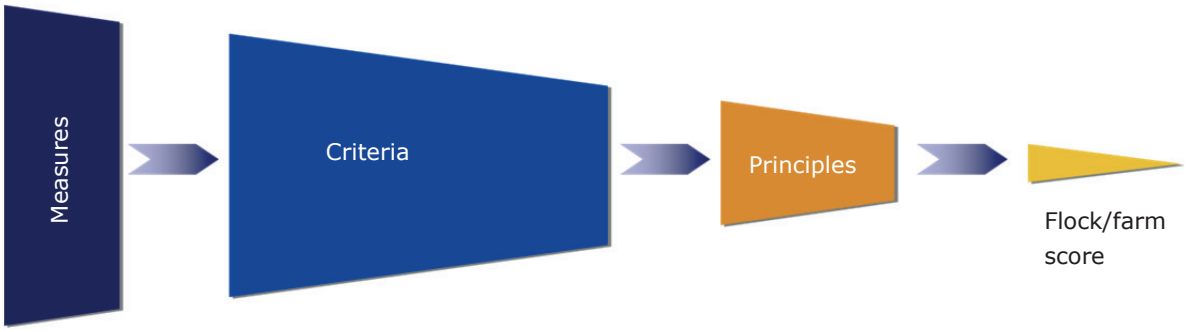

Figure 1. *The Welfare Quality® aggregation model: from measures to final flock/farm score.*

With respect to the WQ broiler assessment protocol, stakeholders expressed their interest in the assessment protocol as a tool to support the management of the farmer and in this way to increase broiler welfare. For broiler chickens, two assessment protocols were developed: one to assess broiler welfare on-farm (see Table 1 in this report) and another protocol to assess broiler welfare during transport and at slaughter (see Table 2 in this report). The aggregation method from measures to criterion, principle and flock scores (Figure 1) has only been developed for the protocol for on-farm assessment of welfare. Therefore, and because of the interest of stakeholders in the protocol to assess on-farm welfare, two studies were carried out in which the broiler assessment protocol was thoroughly studied (De Jong et al., 2011; current report). This resulted in a mass of data regarding the measures in the WQ broiler protocol and the aggregation method, as well as practical assessment experience.

In this chapter, the assessment protocol to measure broiler welfare on-farm is discussed with respect to the following:

(1) Possibilities for simplification. Although the average assessment time is relatively short compared to the assessment protocols for other species, such as dairy cattle or laying hens, a complete assessment on-farm takes 3-4 hours and an additional slaughter plant visit is necessary. Dutch stakeholders emphasized that a reduction in assessment time and/or integration with measures included in legislation will promote adoption of the WQ broiler assessment protocol in practice;

(2)
(3) The relevance of the individual measures with respect to animal welfare. Although an important aspect of the WQ assessment protocols is animal-based measurement (Blokhuis et al., 2010), various management- or resource-based measures have also been included. In addition, measures need to be valid (reflecting the welfare status of the animal);
(4) The aggregation model as proposed by Welfare Quality®. This aggregation model aggregates the individual measures into a single flock or farm score (Figure 1). Relevant questions with respect to application of the WQ broiler protocol in practice are the validity of the model (is the actual variation in on-farm welfare reflected on the level of measures, criteria, principles and final flock scores?) and the relevance of the aggregation method with respect to the use of the WQ broiler protocol as a management-support tool.

3.2 Possibilities for simplification of the WQ broiler protocol

3.2.1 Results from a previous project

In 2011, a large data set of 180 broiler flocks assessed on-farm and 150 flocks assessed at slaughter was used to study possibilities for simplification of the WQ broiler assessment protocol, with the aim to significantly reduce assessment time (De Jong et al., 2011). We will summarise the findings of that study below. Two possible methods of simplification of the WQ broiler protocol were studied:

(1) Use of predictors in on-farm measures. In case of meaningful and significant correlations between on-farm animal-based measures, the value of one measure can be predicted by using the value of another measure. The structure of the WQ assessment protocol (measure-criterion-principle-flock score) remains unchanged;
(2) Use of slaughter plant data to predict on-farm measures. This enables the use of routinely collected data at the plant as predictor of on-farm measures. The structure of the WQ assessment protocol remains unchanged.

The results of the project indicated that there were two possible methods of simplification, based on significant and meaningful correlations:

(1) Predicting gait scores from measures of hock-burn on-farm; a reduction of assessment time on-farm by 1-1.5 h per flock;
(2) Predicting footpad dermatitis and hock burn on-farm from slaughter plant measures of footpad dermatitis and hock burn; predicting cleanliness on-farm from footpad dermatitis and hock burn measures at the slaughter plant; and, predicting gait score on-farm from measures of hock burn at the slaughter plant. Assessment time on-farm will be reduced by 1 h when footpad dermatitis and hock burn at the plant are used to predict the clinical scores on-farm (footpad dermatitis, hock burn and cleanliness). An additional reduction of 1-1.5 h on-farm will be possible when hock burn measures at the plant can predict gait scores on-farm.

Despite a high agreement between the simplified models and the full model on the level of criterion, principle and flock score, it was strongly suggested to re-test the proposed simplification methods in flocks that are widely distributed over the range of different welfare classification categories. Even though we tested the simplification strategies in a large number of flocks with large variation on the level of the individual measures, there was little variation in final flock scores and in scores for some criteria and principles, resulting in a large confidence interval for specificity of the simplified models (De Jong et al., 2011).

3.2.2 Current project: methods

In the current project, the simplification strategies as proposed in the study of De Jong et al. (2011) and summarised in paragraph 3.2.1. were re-tested. In order to increase the probability that flocks were widely distributed over the different WQ classification categories, flocks with standard, fast growing broiler chickens and slower growing strains were measured. Fifty farms participated in the project; 12 farms with slower growing broiler strains and 38 farms with standard, fast growing broilers. For each farm, one house was measured four times. Data of the first two measures (visit 1

and 2, carried out in 2013) were used to re-test the simplification strategies as described above. For the statistical methods to test the simplification strategies we refer to De Jong et al. (2011).

3.2.3 Current project: results

Appendix 3.1. shows the results of the calculation of final flock scores, results of the analysis of correlation between animal-based measures, and the agreement between the full and simplified models on the level of principle and criterion scores that are affected by both simplification strategies.

The basis for simplification is at least a moderate and significant correlation between animal-based measures. During the current study a lower number of flocks were included than in our first study (De Jong et al., 2011), implicating a relatively low sample size, especially with respect to flocks with slower growing broiler strains (N=24 flocks over two subsequent visits). Despite the variation in farming systems, flock scores were not well distributed over the WQ classification categories (Tables 3.1.1. and 3.1.2. in Appendix 3.1). Analysis of correlation showed that, with respect to the measures relevant for simplification, moderate to high and significant correlations were found at the overall level and for flocks with standard broilers only. For flocks with slower growing broilers, significant correlations were only found between measures of footpad dermatitis and hock burn at the plant and the same measures collected on-farm, and between gait score and hock-burn measured at the plant (Table 3.3.3. in Appendix 3.1).

For both simplification strategies scores according to the WQ aggregation method were calculated. Scores for principles and criteria affected by the simplification are shown in Figures 3.3.1 and 3.3.2 in Appendix 3. Figures show that there is a very high correlation between the full and simplified models. However, Tables 3.3.4. to 3.3.7. show that there is a large confidence interval for specificity and sensitivity of both simplified models on the level of principles and criteria affected by simplification. This is caused by the fact that the majority of farms included in the present study were not well distributed over the different WQ classification categories. Thus, similar results were found as compared to our previous study (De Jong et al., 2011).

3.2.4 Discussion

Simplification strategies that were proposed in our previous project (De Jong et al., 2011) were re-tested in the current project using data from 50 flocks, assessed in two visits (with at least one production round in-between two successive visits). In general, results were similar to De Jong et al. (2011): although simplification is promising, flocks were not well distributed over the different WQ classification categories which resulted in large confidence intervals for specificity and sensitivity of the simplified models. Because it had to be decided whether or not to continue with one or both simplified methods for the 3rd and 4th visit of the project, the following arguments had to be taken into consideration.

A positive correlation between footpad dermatitis and hock burn at slaughter, and cleanliness on-farm can be expected because these measures have a common risk factor, i.e. deteriorated litter quality (e.g., de Jong et al., 2014, Shepherd and Fairchild, 2010). In the current project the sample size of flocks with slower growing broilers was relatively small. In our previous project, sample size was larger which resulted in significant correlations between clinical scores on-farm and footpad dermatitis and hock burn scores at slaughter in flocks with slower growing birds, in contrast to the current experiment. In addition, severe forms of footpad dermatitis and hock burn and dirty broilers were uncommon in flocks with slower growing broilers in the current experiment.

Simplification by using slaughter plant measures to predict clinical scores on-farm might lead to a reduction in stress in broiler chickens due to the catching procedure. In addition, this simplification strategy is promising with respect to the integration of the broiler assessment protocol with routinely collected data at the plant.

A correlation between hock-burn and gait score can be expected in situations where broiler chickens decrease activity due to leg problems, thereby increasing the contact with litter and thus increasing the risk of hock burn in case of deteriorated litter quality. However, this also raises the question whether this relationship is always valid, as farmers may be able to keep litter in good condition, while at the same their birds attain a high (worse) gait score. Locomotion is one of the major welfare problems in broiler chickens (EFSA, 2010); it is undesirable that the application of a simplified protocol may lead to under- or over estimation of locomotion problems in a broiler flock. The method for hock burn scoring on the slaughter line seems to be insufficiently distinctive in flocks with standard broilers, as there was no correlation between gait score and hock burn at slaughter and a significant correlation between gait score and hock burn on-farm. Conflicting results were found for flocks with slower growing strains.

Both simplified models showed high agreement with the full model for criteria and principles affected by simplification. However, there was little variation in final flock score (the majority of flocks ended up in the same classification category, i.e. acceptable) and a large confidence interval for specificity and sensitivity of the simplified models was found. Therefore, it was decided to use the analysis of the correlation between individual measures as the basis for a decision on simplification. This led to the decision to apply simplification method 2 and to predict the clinical scores on-farm (cleanliness, footpad dermatitis and hock burn) from footpad dermatitis and hock burn measures at the plant. It was decided not to predict gait score on-farm from hock burn measures at the plant. It should be taken into account that the correlation between footpad dermatitis and hock burn at the plant and cleanliness is lower in flocks with slower growing birds than with fast growing birds, thus, the prediction of cleanliness score in flocks with slower growing birds is expected to be less accurate.

3.2.5 Conclusions

Reduction of assessment time of the WQ broiler protocol will promote application of the assessment protocol in practice. Simplification, and thus time reduction, of the WQ broiler protocol can be achieved by using slaughter plant measures of footpad dermatitis and hock burn to predict the clinical scores (footpad dermatitis, hock burn and cleanliness) on-farm. This simplification strategy reduces on-farm assessment time by approximately one hour. In addition, broilers do not need to be caught for the assessment which reduces stress due to catching procedures. Moreover, by applying this simplification strategy integration with routinely collected data at the slaughter plant is possible. Slaughter plants routinely collect footpad dermatitis and hock burn scores in line with welfare legislation and as indicator of carcass quality.

3.3 Measures

In this chapter, the measures in the WQ broiler assessment protocol are discussed in relation to validity and feasibility.

3.3.1 Principle 1, good feeding

Principle 1, good feeding, is calculated from the criteria absence of prolonged hunger and absence of prolonged thirst. Each of these criteria scores is calculated from one measure (Table 1).

Absence of prolonged thirst is assessed by calculating the number of birds per drinker. Thus far, a feasible animal-based measure for absence of prolonged thirst has not been developed (Vanderhasselt et al., 2013, Vanderhasselt et al., 2014). The recommended number of birds per drinker is based on the Freedom Food certification system and it is yet unclear if these recommendations are scientifically based (Butterworth, pers. comm.). It can be questioned whether or not the number of birds per drinker is a valid measure of absence of prolonged thirst. Additionally , it can also be questioned whether or not the present recommendations included in the protocol are correct.

In conclusion, for the Dutch situation, the current measure for absence of prolonged hunger, i.e. the percentage of emaciated birds at slaughter, is not recorded separately during the *post-mortem* inspection. It is therefore advised to perform a more specified *post-mortem* inspection. The current measure for absence of thirst lacks a scientific basis and its validity can be questioned. Thus, there is a need to develop a valid animal-based measure for absence of prolonged thirst.

3.3.2 Principle 2: good housing

The score for criterion 3, comfort around resting, is based on measures for plumage cleanliness, litter quality and amount of dust. It can be discussed whether or not an additional score for litter quality is required when animal-based measures of plumage cleanliness, footpad dermatitis, hock burn and breast burn are also included in the assessment protocol (Welfare Quality, 2009). These animal-based measures have a common risk factor, i.e. deteriorated litter quality (Allain et al., 2009, Haslam et al., 2007, Hepworth et al., 2010, Shepherd and Fairchild, 2010, Bassler et al., 2013). On the other hand, litter quality scoring is feasible: it is easy to score and because it is combined with the gait scoring it takes only a few minutes per flock.

Measuring cleanliness is valid and feasible. The only disadvantage is that birds have to be caught for inspection. Because cleanliness, footpad dermatitis and hock burns have a common risk factor (deteriorated litter quality), cleanliness scores can be predicted from slaughterhouse scores of footpad dermatitis and hock burn (see chapter 3.2).

The amount of dust in the broiler house is measured by a simple but inaccurate method, i.e. placing a black sheet in the house during the time of observation and scoring the amount of dust on the sheet according to five classes (no dust – lots of dust). Although the amount of dust is important with respect to animal welfare (De Jong et al., 2012) no other feasible (animal-based) measure of dust is available.

The score for criterion 4, thermal comfort, is based on measures of the number of birds panting and huddling. It is likely that under conditions of hot weather, for example, not only the effect of long-term environmental conditions are scored but also the acute, current situation. On the other hand, there are no other feasible animal-based measures of thermal comfort and an estimation of the number of birds showing panting and huddling is easy to perform and does not take long.

The score for criterion 5, ease of movement, is measured by a calculation of the stocking density in the house at the time of assessment. With 'ease of movement' it is meant that animals should have enough space to be able to move around freely; the ability of the bird to move is measured under the criterion 'absence of injuries'. There is currently no other, feasible and valid measure of 'ease of movement' and it is recommended to develop a feasible animal-based measure of ease of movement. It has been suggested that the speed of filling empty spaces in the broiler house might provide an indicator of 'ease of movement' although validation of this measure is necessary. In addition, the maximum stocking density during a production round might be a better indicator than the actual stocking density at the time of visit. An assessment visit is carried out between 5-1 days before slaughter and the stocking density usually increases considerably during these days, especially when thinning is applied and the visit is performed just after thinning.

In conclusion, all measures for the principle 'good housing' are feasible, but the measures need to be reconsidered or improved because of inaccuracy (dust scoring), redundancy (cleanliness, litter quality) or validity (stocking density).

3.3.3 Principle 3: good health

The score for criterion 6, absence of injuries, is based on measures of footpad dermatitis, hock burn, breast burn and lameness. For a discussion on these measures we refer to chapter 2. These measures are valid and feasible, although the assessment of lameness is time consuming and assessors need to

be trained thoroughly. The assessment and description of breast burn according to the WQ broiler assessment protocol (Welfare Quality, 2009) requires improvement. Currently only breast blisters are included, which are hardly observed in broiler chickens, and breast irritation or buttons are not included. It is advised to change the current classification in two classes (0: no breast blisters; 1: evidence of breast blister) into a three point scale: 0: no blisters or irritation; 1: evidence of breast irritation (brown discoloration of the skin); 2: breast blister.

The score for criterion 7, absence of disease, is based on the percentage of mortality and culled birds and the percentage of birds with ascites, dehydration, septicaemia, hepatitis, pericarditis and abscesses using data from slaughterhouse records of rejections. Although there is no doubt on the validity of the measure separate scoring mortality and culls and the various reasons of rejection during *post-mortem* inspection at the plant is not commonly performed[1]. It is preferred to improve the scoring of mortality on Dutch farms, discriminating between culls and birds found dead for all flocks (including flocks housed at stocking densities lower than 39 kg/m^2), and to more specifically score the rejections during *post-mortem* inspection at the plant (see also 3.3.1, emaciation). If this cannot be done, adapted calculations need to be used (see 3.4. aggregation model).

In conclusion, in general the measures under principle 3, good health, are valid and feasible. For countries where separate scoring of culls and birds found dead, and separate scoring of the various reasons for rejections at *post-mortem* inspection are uncommon, it is advised to perform a more specified *post-mortem* inspection or on-farm record keeping.

3.3.4 Principle 4: appropriate behaviour

There is yet no measure for criterion 9, expression of social behaviours, in the WQ broiler assessment protocol. In the WQ laying hen assessment protocol aggressive pecking and its consequences (plumage and comb damage) are measured under this criterion. These measures are most likely not included in the broiler protocol because aggressive behaviour is rare in broiler chickens approaching slaughter age. However, if this behaviour occurs in a flock it might be important to record even it is observed at low frequencies. For example, scan sampling of social behaviour can be performed within a reasonable time span after a short training and could provide a good impression of social behaviours.

Cover on the outdoor range and the percentage of birds in the outdoor range are measured under criterion 10, expression of other behaviours. It is unclear why these resource-based measures have been selected for this criterion. In addition, it is unclear why only outdoor ranges are judged positively and covered verandas or the presence of environmental enrichment are not judged positively. Both verandas and enrichments stimulate the performance of natural behaviour, such as foraging, explorative pecking, perching and dust bathing. It is strongly advised to further develop animal-based measures for this criterion. For example, scan sampling can be used to form an impression of the extent to which (natural) behaviours are carried out, such as foraging and dust bathing. In addition, until an animal-based measure has been developed, it is advised to provide a positive assessment of systems using covered verandas and enrichments.

Scores in the touch test are used to calculate a score for criterion 11, good human-animal relationship. In the touch test the observer approaches a group of birds, squats for ten seconds, counts the number of birds at arm's length and calculates the number of birds that can be touched. However, in a previous study it was already discussed whether the touch test is a good indicator of the human animal relationship, i.e. whether it actually measures fear for humans. The response appeared to be positively correlated with the locomotion score of the birds in a flock and the stocking density (De Jong et al., 2011), despite the correction for stocking density when calculating the score for this criterion.

[1] Because of the Broiler Directive, for flocks kept at stocking densities higher than 39 kg/m^2 the number of birds culled and found dead need to be registered separately. However, this is not common practice in flocks kept at lower stocking densities (<33 kg/m^2).

In addition, observers indicated that flocks that showed low responses in the touch test, thus were scored as being fearful for humans, appeared not to be fearful in the opinion of the observers. This might be due to the fact that these birds initially moved away from the observer, which was scored, but subsequently approached the observer, which was not scored in the test. It is strongly advised to validate the touch test for broiler chickens as this had not been done in the development of the WQ broiler assessment protocol (Niebuhr, pers. comm.) and/or to develop a new test to measure the human-animal relationship. For example, an approach test, which does not measure the initial natural response (moving away) but the latency to approach an unknown person, might give a good indication of fear for humans. Such a test requires validation.

The Qualitative Behaviour Assessment (QBA test) is used for scoring criterion 12, positive emotional state. The QBA has been developed for pigs (Wemelsfelder et al., 2000). Initially, (a group of) animals were observed during a certain time period, after which the behaviour of the animals was described. After development and validation in pigs the method has been developed further and applied in various animal species, i.e. horses, goats and dairy cattle. For the WQ assessment protocols for the various species it was chosen to use a list of standard terms to describe the behaviour of the animals instead of a free choice by the observer (see e.g. Welfare Quality, 2009). The application of the QBA test in the WQ assessment protocol has led to discussion on the validity of the tool. First, the QBA has never been validated for broiler chickens (Wemelsfelder, pers. comm.). Second, for a reliable scoring of a group of broiler chickens, observers need to have sufficient knowledge of the species and its behaviour. Scoring the behaviour of a broiler flock using terms such as 'inquisitive', 'bored', 'positively occupied', etc. is much more difficult for people not having experience in the observation of broiler flocks, in contrast to using e.g. scan sampling to score behaviour of the birds (De Jong, pers. comm.). Third, it is reasonable to assume that a QBA can be scored better for large animals kept in small groups, showing postures, facial expressions and vocalisations, than chickens (broilers and laying hens) housed in groups of several thousands of birds. And finally, when communicating the test results to farmers, the test results are difficult to explain, and calculation of the QBA flock behaviour scores is complex.

It is strongly advised to validate the QBA in broiler chickens and to test inter- and intra-observer reliability. Or, alternatively, an alternative measure needs to be developed for criterion 12, positive emotional state. One suggestion is to use scan sampling to score birds for positive behaviours such as dust bathing or exploration, or those that do not display any negative behaviours such as fearfulness or aggressive pecking.

Because the principle 'appropriate behaviour' is an important aspect of animal welfare, we advise that to work should be done on the development and improvement of measures to include principle 4 (see also discussion below).

In conclusion, although the ability of the broiler chicken to perform its natural behaviour is an important aspect of its welfare, the measures under the principle 'appropriate behaviour' do not meet requirements for validity and feasibility. In case of the application of resource-based measures (criterion 10, expression of other behaviours), as long as no valid and feasible animal-based measures are available, it is advised to include other resource-based measures such as the presence of enrichments or a covered veranda. It is strongly advised to develop valid animal-based measures for all criteria and/or to perform a validation study with respect to the touch test and QBA.

3.4 Aggregation model

Welfare Quality® applies an aggregation model to calculate a final flock score from individual measures collected on-farm or at slaughter (Welfare Quality, 2009) (see Figure 1). This aggregation model will be discussed in relation to the measures in broiler chickens and its sensitivity to reflect broiler welfare.

3.4.1 Aggregation model: adaptations

Because in the Dutch situation a few measures as described in the WQ broiler assessment protocol could not be collected, modifications to the aggregation model were required in order to calculate the principle, criterion and final flock scores. Without these modifications it would not have been possible to calculate these scores, as the WQ aggregation model does not allow missing values for measures.

The modifications of the aggregation model were as follows:

1. As no figures for the percentage of emaciation could be collected, no scores for criterion 1, absence of prolonged hunger, could be calculated. Principle 1 was therefore only based on the scores for criterion 2, absence of prolonged thirst.
2. Because various farmers did not separately register the number of culled birds and the number of birds found dead, the score for criterion 7 (absence of disease) could not be calculated. In our previous study a new expert opinion was carried out to develop a new spline function and Choquet integral for criterion 7, based on total mortality figures only (De Jong et al., 2011). In addition to the mortality figures, rejection figures were used to calculate the score for criterion 7. According to the WQ protocol, separate figures for the percentage of birds with ascites, dehydration, septicaemia, hepatitis, pericarditis and abscesses should be collected (usually part of the *post-mortem* inspection figures) (Welfare Quality, 2009). In the Dutch situation, only total rejection percentage is available. Therefore a new expert opinion was performed for the total rejection percentage and a new spline function and choquet integral were calculated as described in De Jong et al. (2011), to enable calculation of a score for criterion 7 (absence of disease).

In conclusion, it is strongly recommended to provide alternatives, as mentioned above, to calculate an aggregation score in case some measures cannot be collected. In addition, it is recommended to develop guidelines to calculate principle, criterion and flock scores in case of a missing value for a measure, because the current model does not allow missing values in the aggregation model.

3.4.2 Aggregation model: weighting and calculation of final flock scores

When more than one measure is used to calculate a criterion score, expert opinion was sought to construct the aggregation model for the criterion score. Experts were consulted to weigh the different measures to build a score for a criterion. Similarly, each principle score is calculated from more than one criterion score. Expert opinions have led to weighting for each criterion score. Experts in the field of broiler welfare were consulted to weigh the measures to build a criterion score (Van Reenen, pers. comm.). Experts and stakeholders were consulted to build the principle scores from the criteria scores, and to build the flock score from the principle scores (http://www1.clermont.inra.fr/wq/index.php?id=farms#criteria). Where the aggregation from measures to criterion scores is species specific, a general approach was used for all species to aggregate criterion scores into principle scores and principle scores into flock scores (Van Reenen, pers. comm.).

An important aspect of the WQ aggregation model is that it does not allow compensation on the level of measures and criteria. Thus, for example, a flock having a low score for one criterion cannot be compensated by a high score for another criterion when calculating the principle score.

Below, the aggregation and weighting is discussed in relation to the validity of aggregation in the WQ broiler protocol and illustrated with some examples. More information on the aggregation model can be found at http://www1.clermont.inra.fr/wq/index.php?id=farms#criteria and in Keeling (2009).

3.4.3 Principle 1: good feeding

For the calculation of principle 1, good feeding, absence of prolonged thirst is considered to be more important than absence of prolonged hunger, thus, measures of the number of birds per type of drinker is considered more important than the percentage of emaciated birds. In case of the

impossibility of separate scoring of emaciation (Dutch flocks) the principle score is similar to the criterion score for absence of prolonged thirst (measure of number of birds per drinker). Above, we discussed the validity of the measure of number of birds per drinker as indicator of thirst. Here, we illustrate that at least in the Dutch situation it leads to principle scores that are open to debate.

It was observed that flocks with slower growing broiler chickens usually have more birds per drinker than flocks with standard, fast growing broiler chickens. This results in a lower percentage of drinker compliance (% compliance = number of recommended birds per drinker/number of birds per drinker * 100). Figure 2 shows the average compliance and criterion 2 scores for flocks with standard and slower growing strains.

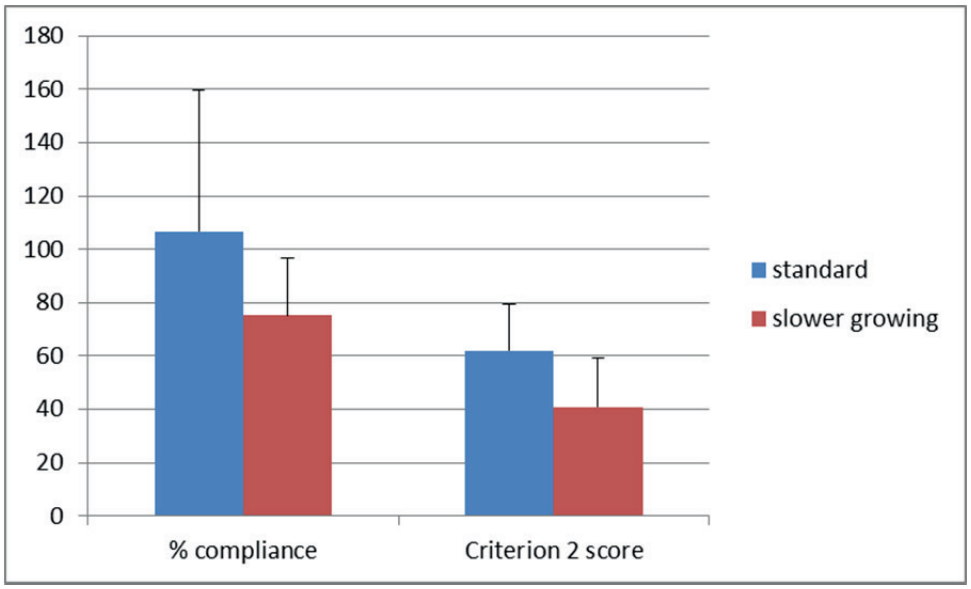

Figure 2. *Average percentage of drinker compliance (left) and the average scores for criterion 2 (absence of prolonged thirst) for flocks with standard broilers and flocks with slower growing broiler strains assessed in visit 3.*

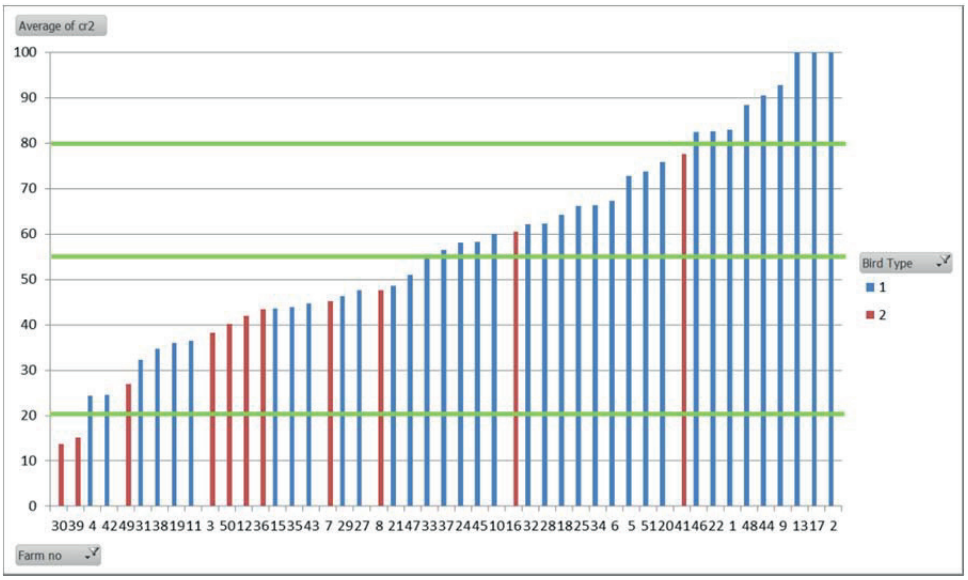

Figure 3. *Score for criterion 2, per flock assessed in visit 3. As there is no measure for criterion 1, this score is similar to the principle 1 score. Thresholds according to the WQ assessment protocol that are considered for final flock classification are indicated with green lines (20, 55 and 80 points). Red bars are flocks with slower growing broiler strains, blue bars are flocks with standard broiler strains.*

Figure 3 shows the criterion scores for each flock. Scores for principle 1 are, at least in the Dutch situation, similar to scores for criterion 2. These results are open to debate because (1) the results indicate that flocks with slower growing strains have a low (bad) score for absence of prolonged thirst, and (2) the results indicate that there is a large variability in the degree in which birds suffer from prolonged thirst. However, in practice, there are no indications that flocks with slower growing strains are more thirsty as compared to flocks with standard broilers or that there is such a large variability in the degree to which the assessed flocks suffer from thirst. Especially the low scores for the flocks with slower growing birds, of which the validity is debated, affect the probability that a flock will receive a high welfare classification. For example, flocks that have a principle score lower than 20 can never get the final score 'enhanced', even if the scores for all other principles are high[2].

In conclusion, criterion 2/principle 1 scores do not seem to reflect actual differences in levels of chronic thirst and thus in welfare between broiler flocks.

3.4.4 Principle 2: good housing

Within criterion 3, the measure of litter quality is considered more important than cleanliness, which is considered more important than dust. Within criterion 4 (thermal comfort) the percentage of birds showing panting or huddling is combined as a single measure (e.g., more than 50% of the birds, but less than 100%, pant or huddle). Criterion 5 (ease of movement) is based on stocking density only.

In the aggregation of the scores for the criteria into the score for principle 2, criterion 5, ease of movement, is considered to be more important than criterion 3, comfort around resting, which is considered more important than criterion 4, thermal comfort. In summary, stocking density has the largest effect on the principle score. Animal-based measures such as cleanliness and panting/huddling have a relatively low weighting.

This might result in principle scores that seem to be relatively less sensitive to flock variability in animal-based measures. Stocking density at the time of assessment does not show large variation within farming systems because farmers aim to keep their birds at the highest density allowed for a specific farming system for economic reasons. A few examples are shown in Table 4. Criterion 5 scores are based on stocking density and only differ between flocks with standard or slower growing broiler chickens. Flocks 11 and 12 show similar scores for criterion 3 and 4 but a large difference in criterion score 5, resulting in a large difference in principle 2 score. Flocks 47 and 48 show similar scores for criterion 4 and 5 but differ in criterion 3 score, which has little effect on principle 2 score. Flocks 49 and 50 show similar scores for criterion 5 but large differences in scores for criterion 3 and 4, but this results in a relatively small score for principle 2.

[2] A flock is classified 'excellent' if it scores more than 55 on all principles and more than 80 on two principles; it is classified 'enhanced' if it scores more than 20 on all principles and more than 55 on two of them; it is classified 'acceptable' if it scores more than 10 on all principles and more than 20 on three of them. Flocks that do not reach these standards are 'not classified' (Welfare Quality, 2009).

Table 4

Examples of scores for criteria 3, 4 and 5, and the score for principle 2 for a few flocks (assessments in visit 3).

Flock	Bird type	Principle 2 score[1]	Criterion 3 score[1]	Criterion 4 score[1]	Criterion 5 score[1]
11	Standard	46.776156	65.2753	69	37.94413
12	Slower growing	65.47516	67.25721	69	64.42392
47	Standard	47.274287	65.24487	69	38.6748
48	Standard	41.510085	44.40486	69	36.65041
49	Slower growing	61.394926	59.70793	69	62.42698
50	Slower growing	70.104781	70.90903	100	65.53111

[1] Criterion 3: comfort around resting; Criterion 4: thermal comfort; Criterion 5: ease of movement; Principle 2: good housing.

In conclusion, scores for principle 2 are largely affected by scores for criterion 5, which is a disputed measure. The animal-based measures for criterion 3 and 5 had little effect on the principle scores. It is therefore recommended to reconsider the aggregation method for principle 2.

3.4.5 Principle 3, good health

Within principle 3 'good health' criterion 7 (absence of disease) is considered more important than criterion 6 (absence of injuries), which is in turn considered more important than criterion 8 (absence of pain induced by management procedures, which is not applicable to broiler chickens). Thus for broiler chickens, the principle 3 score is based on scores for criteria 6 and 7. Criterion 6 scores are based on measures of gait score, hock burn, breast blisters and footpad dermatitis which are weighted more or less equally in the criterion score.

Figure 4 shows examples of scores for principle 4, criterion 6 and criterion 7. In general, flocks with slower growing broiler chickens had low prevalence of lameness, footpad dermatitis, hock burn and breast blisters. This is reflected in the scores for criterion 6, these being highest for flocks with slower growing birds. However, figure 5 also shows that, in general, all scores are moderate despite the low prevalence of contact dermatitis and lameness in slower growing broiler strains. This indicates that there is little tolerance for prevalence of contact dermatitis and lameness.

Figure 6 shows that much higher scores were found for criterion 7. This criterion received a higher weight in the calculation of the score for principle 3, which is illustrated in figure 4; flocks with high (good) scores for criterion 7 received a relatively high score for criterion 5. See for example flock 1 and 2, that received a low score for criterion 6, a high score for criterion 7, and a relatively high score for principle 3.

In conclusion, it is recommended to reconsider the aggregation method for criterion 6 because there appears to be little tolerance for contact dermatitis and lameness scores. A positive note is that for principle 3 between flock variation in measure scores are reflected in criterion scores, and variation in criterion scores are reflected in principle scores.

Figure 4. *Examples of scores for principle 3 (good health, upper graph), criterion 6 (absence of injuries, middle graph) and criterion 7 (absence of disease, lower graph). Assessments were carried out in visit 3. Thresholds according to the WQ assessment protocol that are considered for final flock classification are indicated with green lines (20, 55 and 80 points). Red bars are flocks with slower growing broiler strains, blue bars are flocks with standard broiler strains.*

3.4.6 Principle 4, appropriate behaviour

Criterion 9, expression of social behaviour, is considered least important, but there is currently no measure for this criterion in the WQ broiler protocol. Criterion 11, good animal-human relationship is considered more important than criterion 9. Criteria 10 and 12 (expression of other behaviours and positive emotional state) are considered slightly more important than criterion 9. As discussed in 3.3 the validity of the measures for these criteria is open to debate. Because of a similar and low score for criterion 10 for all indoor farming systems (expression of other behaviour, measured by presence of an outdoor range and the percentage of birds observed in the range) all flocks received a low score for principle 4 (appropriate behaviour), despite a wide variation in scores for criterion 11 (human-animal relationship) and 12 (positive emotional state). This is illustrated in Figures 5 and 6.

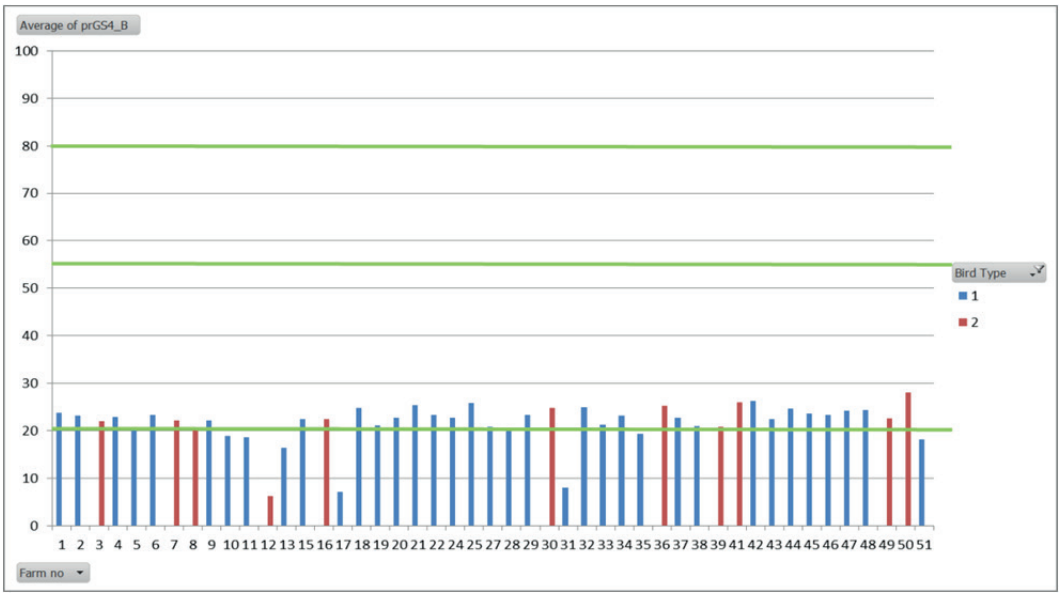

Figure 5. *Scores for principle 4 (appropriate behaviour) in visit 3. Thresholds according to the WQ assessment protocol that are considered for final flock classification are indicated with green lines (20, 55 and 80 points). Red bars are flocks with slower growing broiler strains, blue bars are flocks with standard broiler strains.*

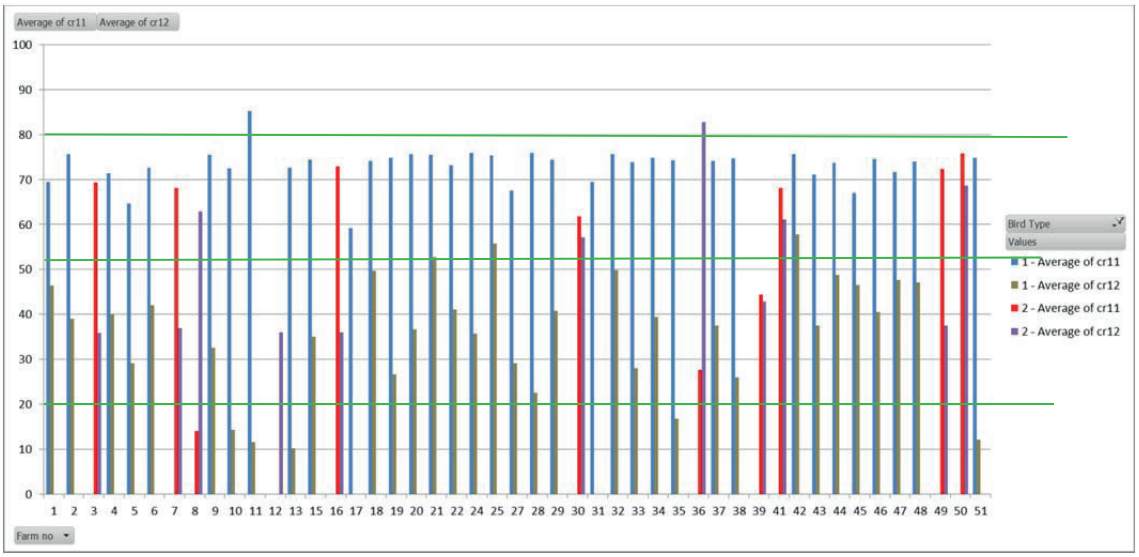

Figure 6. *Scores for criterion 11 (human-animal relationship) and criterion 12 (positive emotional state) in visit 3. Thresholds according to the WQ assessment protocol that are considered for final flock classification are indicated with green lines (20, 55 and 80 points). Blue and green bars are the scores for flocks with fast growing birds, red and purple bars are the scores for flocks with slower growing birds.*

When comparing figures 5 and 6 it becomes clear that despite a wide variation in scores for criteria 11 and 12 this between-flock variation is not reflected in the principle scores. This is explained by the fact that all flocks received the lowest score for criterion 10 (expression of other behaviours measured by the presence and percentage of birds in the outdoor range). These examples show that because of the relatively high weighting of a disputed measure (outdoor range) flocks cannot receive high scores for principle 4. An implication is that the probability to receive the classification 'enhanced' is low and that it will be impossible to receive the final classification category 'excellent' for all indoor farming systems.

Figure 7 further illustrates the above. In this figure two scores for principle 4 are compared, one according to the original WQ broiler assessment protocol, and one in which systems with a covered outdoor range that was accessible at the time of visit, received a score for the number of birds in the covered outdoor range. This resulted in much higher principle score for these farms (purple bars in Figure 7).

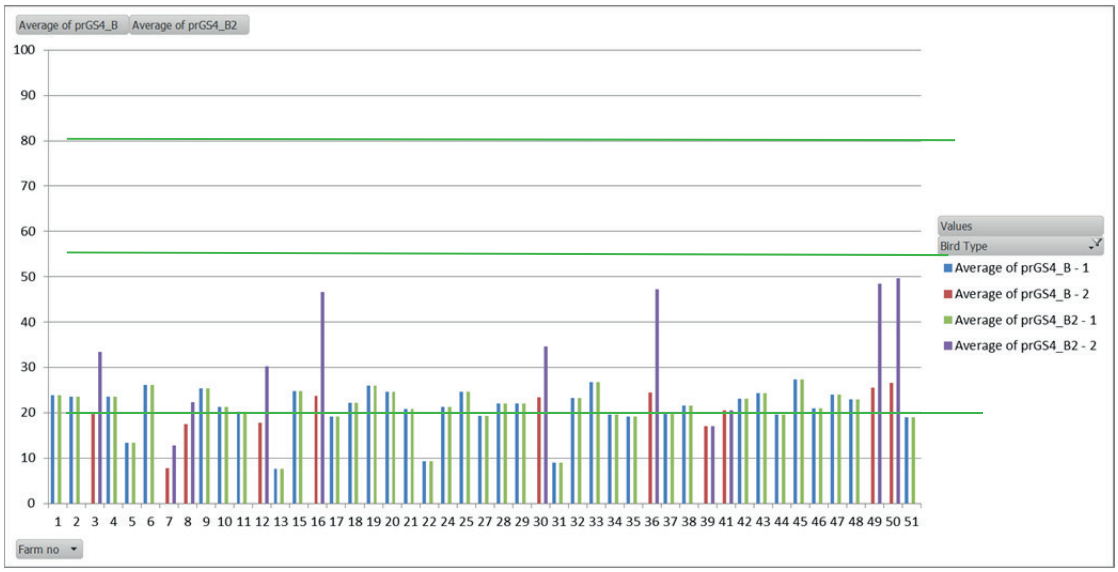

Figure 7. *Scores for principle 4 according to the WQ broiler assessment protocol (prGS4_B) and an adapted measure for criterion 9 where the percentage of birds in the covered outdoor was also included (prGS4_B2). Bird type 1 were flocks with fast growing broilers, bird type 2 were flocks with slower growing broiler strains. Data collected in visit 4. Thresholds according to the WQ assessment protocol that are considered for final flock classification are indicated with green lines (20, 55 and 80 points).*

In conclusion, principle 4 score is largely affected by the criterion 10 score. Because the measures are disputed, it is recommended to reconsider the aggregation method after valid measures have been developed for criteria 9 to 12.

3.5 Recommendations and implementation in practice

In chapter 3 we critically reviewed the various aspects of the WQ broiler assessment protocol. Although there is a need for improvement in relation to the validity of the individual measures and the aggregation method, parts of the WQ broiler assessment protocol are well developed with valid measures.

An important aspect of implementation in practice was the possibility to reduce assessment time and/or the integration with existing measures at the slaughter plant. The current study and our previous study (De Jong et al., 2011) indicate that simplification, using slaughter plant measures of footpad dermatitis and hock burn to predict cleanliness on-farm, is possible and reduces assessment time on-farm with 1-1.5 h.

The following animal-based measures have been shown to be feasible and valid: footpad dermatitis, hock burn, breast blister, lameness, mortality and culls, rejection figures, cleanliness, panting and

huddling. It is advised to improve measures of rejection and mortality by more specifically scoring rejections and separate scoring of culls for all farming systems. If that is possible then an animal-based measure will be available for criterion 1 (emaciated birds). If this is impossible, modified calculations have been presented to estimate criterion scores.

The following animal-based measures need to be improved or validated: breast blisters (also taking into account breast irritation), Qualitative Behaviour Assessment, touch test.

With respect to the resource-based or management-based measures, it is advised to develop animal-based measures where possible, but especially with respect to criterion 10, expression of other behaviours, criterion 5, ease of movement and criterion 2 (absence of thirst). For measurement of dust, it is preferred to develop an animal-based measure or to develop a more accurate scoring method, but as it is undesirable to use (expensive) tools for assessment it can be questioned if there are any feasible alternatives.

In addition, because appropriate behaviour is an important welfare aspect it is strongly recommended to give priority to the development of valid animal-based measures for this principle as there is currently no animal-based measure available.

Before parts of the WQ broiler assessment protocol can be implemented in practice as a management tool for farmers or advisors to improve on-farm welfare, there are some other considerations that need to be taken into account.

During the current study farms were visited four times. Especially farmers with slower growing broiler strains indicated that the visit frequency could be reduced. In general, we got the impression that there was a higher need for feedback on farms having standard broilers than for those with slower growing broiler strains. Because farms with standard broilers run more production rounds per year, visit frequency should be adapted to the farming system. But we also got the impression that the variation between farms and within farms between successive flocks was larger in farms with standard broilers than for farms with slower growing broiler strains, except for systems with an outdoor range where probably season has a relatively large effect on some welfare indicators.

Another consideration with respect to visit frequency might be the use of iceberg indicators. If, for instance, slaughterhouse measures of contact dermatitis and rejections can be used as iceberg indicator, visits might be limited to farms where welfare is at risk. The use of iceberg indicators merits further study.

During the current project our feedback to farmers was on the level of the individual measures. As discussed in chapter 3.4, the WQ aggregation method for some criteria and principles seemed to be insensitive (not reflecting actual differences in welfare between flocks). In addition, individual measures have greater value for a farmer than integrated scores for criteria and principles. In the time span during which our visits were carried out many farmers made changes to their management. One reason for management change was to improve the footpad dermatitis scores. Management tools are available to reduce the prevalence of footpad dermatitis. However, for example with respect to lameness, management has relatively little effect (EFSA, 2010). If (parts of) the WQ broiler assessment protocol is to be implemented in practice as a tool for management, farmers need to be facilitated with sufficient tools for improvement.

The current study and our previous work has been focused on the WQ broiler assessment tool to improve on-farm welfare. However, welfare during catching, transport, unloading, stunning and killing is also important. WQ has developed an assessment protocol for the end-of-life stage but this protocol requires further developed (Table 2).

Besides the application as management support tool, the WQ broiler assessment protocol can be used to compare farming systems with respect to animal welfare.

In conclusion, parts of the broiler assessment protocol are suitable as management tool for farmers to improve the welfare of broiler chickens on-farm. It is advised to start with the feasible and valid animal-based measures. At the same time, animal-based measures should be developed where currently information is lacking or invalid so that in the future more welfare aspects can be included in the broiler assessment protocol. In addition, if overall flock classification is required, the aggregation method will have to be improved so that actual differences in welfare level between flocks are reflected in the final flock classifications.

3.6 Literature

Allain, V., Mirabito, L., Arnould, C., Colas, M., Le Bouquin, S., Lupo, C. & 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**,** 407-417.

Bassler, A. W., Arnould, C., Butterworth, A., Colin, L., De Jong, I. C., Ferrante, V., Ferrari, P., Haslam, S., Wemelsfelder, F. & Blokhuis, H. J. 2013. Potential risk factors associated with contact dermatitis, lameness, negative emotional state, and fear of humans in broiler chicken flocks. *Poultry Science,* 92**,** 2811-2826.

Blokhuis, H. J., Veissier, I., Miele, M. & Jones, B. 2010. The Welfare Quality (R) project and beyond: Safeguarding farm animal well-being. *Acta Agriculturae Scandinavica Section a-Animal Science,* 60**,** 129-140.

de Jong, I. C., Gunnink, H. & van Harn, J. 2014. Wet litter not only induces footpad dermatitis but also reduces overall welfare, technical performance, and carcass yield in broiler chickens. *Journal of Applied Poultry Research,* 23**,** 51-58.

De Jong, I. C., Perez Moya, T., Gunnink, H., Van den Heuvel, H., Hindle, V., Mul, M. & Van Reenen, C. G. 2011. Symplifying the Welfare Quality assessment protocol for broilers. Report 533. Wageningen UR Livestock Research, Lelystad.

De Jong I, Berg C, Butterworth A and Estevez I 2012c. Report updating the EFSA opinion on the Welfare of Broilers and Broiler Breeders – Part A – The Welfare of Chickens Kept for Meat Production. European Food Safety Authority, Parma, Italy.

EFSA. 2010. Scientific opinion on the influence of genetic parameters on the welfare and the resistance to stress of commercial broilers. *EFSA Journal,* 8**,** 1666.

Haslam, S. M., Knowles, T. G., Brown, S. N., Wilkins, L. J., Kestin, S. C., Warriss, P. D. & Nicol, C. J. 2007. Factors affecting the prevalence of foot pad dermatitis, hock burn and breast burn in broiler chicken. *British Poultry Science,* 48**,** 264-275.

Hepworth, P. J., Nefedov, A. V., Muchnik, I. B. & Morgan, K. L. 2010. Early warning indicators for hock burn in broiler flocks. *Avian Pathology,* 39**,** 405-409.

Keeling, L.J., 2009. An overview of the Development of the Welfare Quality® assessment systems. Welfare Quality Reports no. 12, Welfare Quality, Lelystad, The Netherlands.

Welfare Quality. 2009. *The Welfare Quality assessment protocol for poultry (broilers, laying hens)*. The Welfare Quality Consortium, Lelystad.

Shepherd, E. M. & Fairchild, B. D. 2010. Footpad dermatitis in poultry. *Poultry Science,* 89**,** 2043-2051.

Vanderhasselt, R. F., Buijs, S., Sprenger, M., Goethals, K., Willemsen, H., Duchateau, L. & Tuyttens, F. A. M. 2013. Dehydration indicators for broiler chickens at slaughter. *Poultry Science,* 92**,** 612-619.

Vanderhasselt, R. F., Goethals, K., Buijs, S., Federici, J. F., Sans, E. C. O., Molento, C. F. M., Duchateau, L. & Tuyttens, F. A. M. 2014. Performance of an animal-based test of thirst in commercial broiler chicken farms. *Poultry Science,* 93**,** 1327-1336.

Wemelsfelder, F., Hunter, E. A., Mendl, M. T. & 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**,** 193-215.

Appendix 3.1. Results of analysis of simplification of the WQ broiler assessment protocol

Table 3.1.1.

Distribution of the flocks assessed in visit 1 and visit 2 over the WQ categories, calculated with the full WQ model as described in (Welfare Quality, 2009).

WQ category	# flocks visit 1	# flocks visit 2
Not classified	10	10
Acceptable	37	39
Enhanced	0	0
Excellent	0	0
NA[1]	3	0

[1] NA: final flock score could not be calculated due to missing data

Table 3.2.2.

Distribution of the flocks assessed in visit 1 and 2 over the WQ categories and over the broiler farming systems included in the project, calculated with the full WQ model as described in (Quality, 2009).

Visit #	Farming system	WQ category		
		Not Classified	**Acceptable**	**NA[1]**
Visit 1	Standard (fast growing broilers)	7	28	3
	'Beter Leven 1 ster' (slower growing broiler strains)	3	7	0
	Organic (slower growing broiler strains)	0	2	0
Visit 2	Standard (fast growing broilers)	6	31	0
	'Beter Leven 1 ster' (slower growing broiler strains)	4	6	0
	Organic (slower growing broiler strains)	0	2	0

[1] NA: final flock score could not be calculated due to missing data

Table 3.3.3.

Spearman rank correlations between animal-based measures relevant for further simplification of the WQ broiler assessment protocol. Correlations were calculated for all flocks (overall), and for flocks with standard broilers and slower growing broilers separately. Correlations were calculated for data collected in the first two farm visits (N=100 broiler flocks; N=76 standard flocks; N=24 flocks with slower growing broilers).

Measures	r_{sp} overall	r_{sp} fast growing	r_{sp} slower growing
Gait score ≥ 3 and % HB≥ 3(F)	0.649†	0.444†	*
% HB1 (S) and % HB ≥ 3(F)	0.653†	0.476†	0.346‡
% FPD ≥ 3 (S) and % FPD ≥ 3 (F)	0.831†	0.788†	0.636†
% FPD ≥ 3 (S) and % cl ≥ 2 (F)	0.388†	0.401†	*
% HB 1(S) and % cl ≥ 2 (F)	0.397†	0.324†	*
% gait ≥ 3 and HB 1 (S)	0.665†	* (0.283†)	0.395‡

† Significant correlation, P<0.05

‡ Trend, P<0.10

* low correlation r_{sp}<0.3 and not significant

() significant (P<0.05) but low correlation r_{sp}<0.3

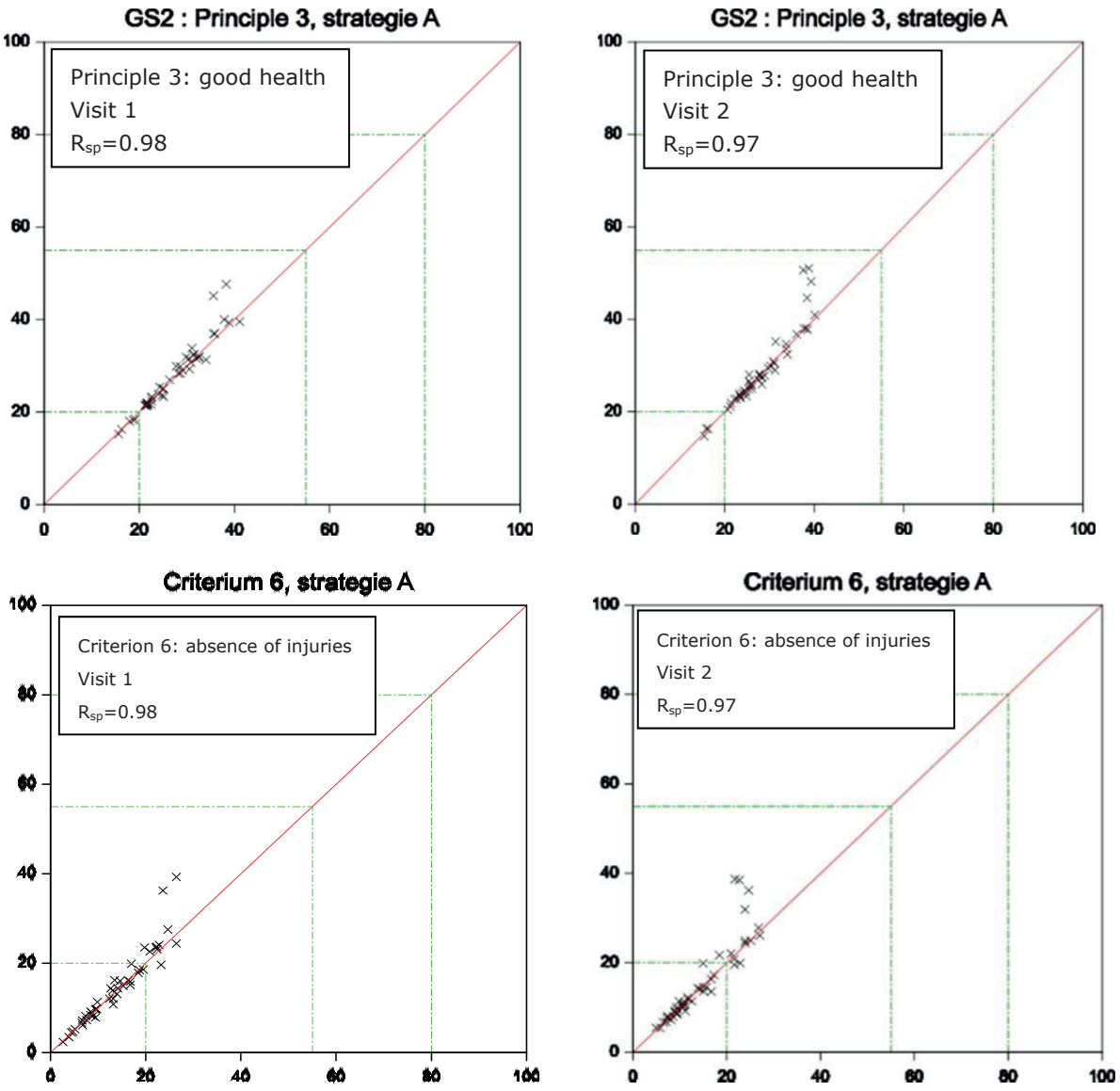

Figure 3.1.1. *Correlation between the full WQ model, and simplified model 1: predicting gait score on-farm from hock-burn on-farm. The figures show the correlation between the full model on the Y-axis and the simplified model on the X-axis, for the principles and criteria affected by the simplification. Spearman rank correlations between scores of the full and simplified model are indicated in the figures (r_{sp}).*

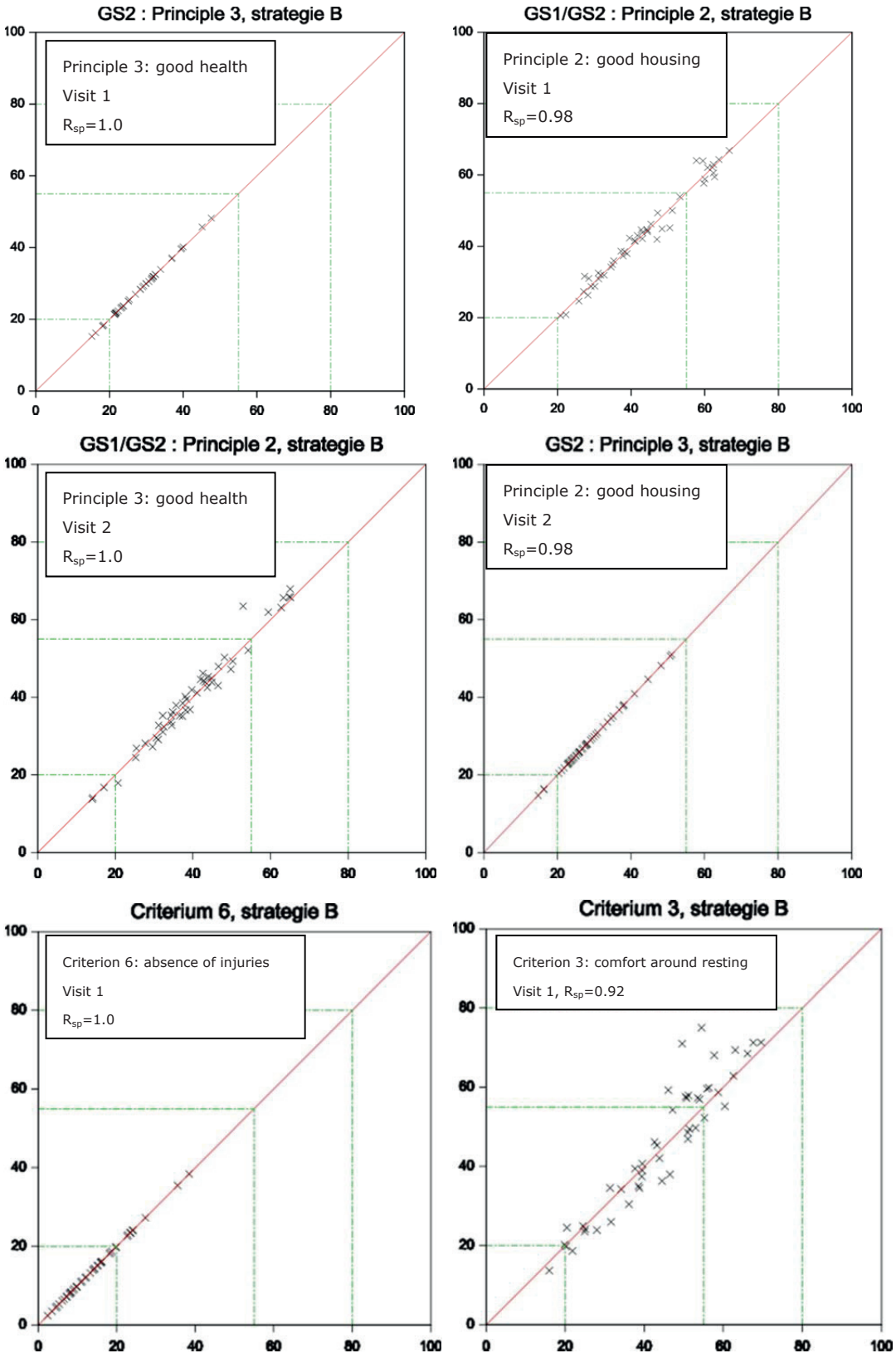

Figure 3. *For legend, see next page.*

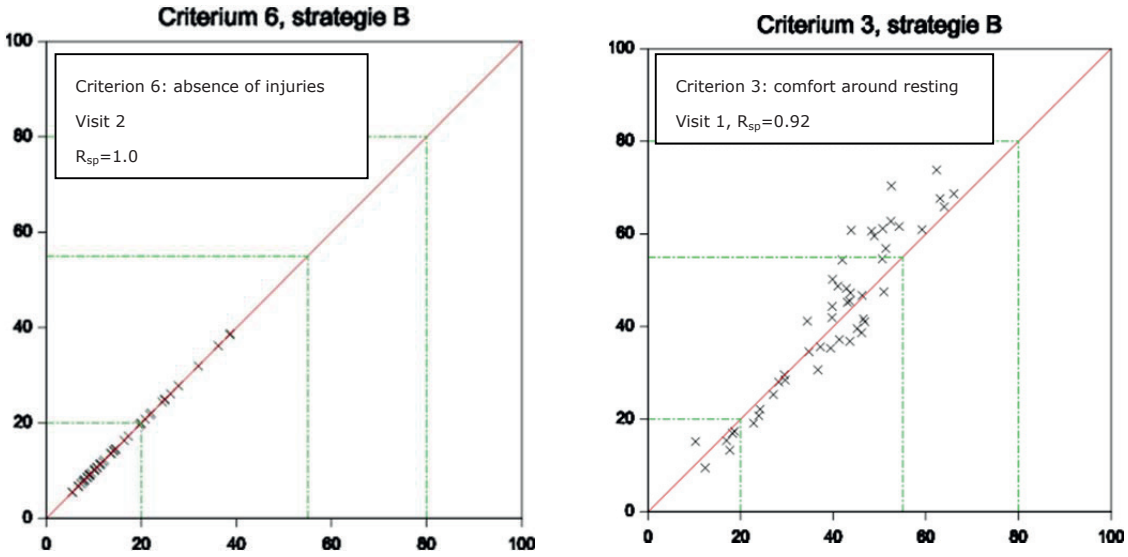

Figure 3.1.2. *Correlation between the full WQ model, and simplified model 2: prediction of clinical scores on-farm by slaughter plant measures of footpad dermatitis and hock burn, and prediction of gait scores on-farm from hock burn measured at the plant. The figures show the correlation between the full model on the Y-axis and the simplified model on the X-axis, for the principles and criteria affected by the simplification. Spearman rank correlations between scores of the full and simplified model are indicated in the figures (r_{sp}).*

Table 3.1.4.

Comparison of principle scores using the full model or simplification strategy model 1 (predicting gait scores from measures of hock burn on-farm), indicated as distribution of flocks across final flock scores as defined by Welfare Quality® including estimates for agreement between methods (Est), and the lower (Low) and upper (Upp) confidence limits within classification groups for equality (% eq);specificity (%sp), sensitivity (%se), Table A shows the result for visit 1, table B shows the results for visit 2.

Table A. Visit 1.

Principle 3		20.0			55.0			80.0		
		90% Conf.interval			90% Conf.interval			90% Conf.interval		
		est.	lower	upper	est.	lower	upper	est.	lower	upper
	%equal	97.9	90.3	99.9	100.0	93.8	100.0	100.0	93.8	100.0
	%se	100.0	22.4	100.0	100.0	93.8	100.0	100.0	93.8	100.0
Rsp=0.98	%sp	97.8	89.9	99.9						

Table B. Visit 2.

Principle 3		20.0			55.0			80.0		
		90% Conf.interval			90% Conf.interval			90% Conf.interval		
		est.	lower	upper	est.	lower	upper	est.	lower	upper
	%equal	100.0	94.1	100.0	100.0	94.1	100.0	100.0	94.1	100.0
	%se				100.0	94.1	100.0	100.0	94.1	100.0
Rsp=0.97	%sp	100.0	94.1	100.0						

Table 3.1.5.

Comparison of principle scores using the full model or simplification strategy model 2 (predicting clinical scores and gait score on-farm from slaughterhouse measures of footpad dermatitis and hock burn), indicated as distribution of flocks across final flock scores as defined by Welfare Quality® including estimates for agreement between methods (Est), and the lower (Low) and upper (Upp) confidence limits within classification groups for equality (% eq);specificity (%sp), sensitivity (%se), Table A shows the result for visit 1, table B shows the results for visit 2.

Table A. Visit 1.

Principle 3		20.0			55.0			80.0		
		90% Conf.interval			90% Conf.interval			90% Conf.interval		
		est.	lower	upper	est.	lower	upper	est.	lower	upper
	%equal	100.0	93.8	100.0	100.0	93.8	100.0	100.0	93.8	100.0
	%se	100.0	22.4	100.0	100.0	93.8	100.0	100.0	93.8	100.0
Rsp=1.00	%sp	100.0	93.6	100.0						

Principle 2		20.0			55.0			80.0		
		90% Conf.interval			90% Conf.interval			90% Conf.interval		
		est.	lower	upper	est.	lower	upper	est.	lower	upper
	%equal	100.0	94.0	100.0	100.0	94.0	100.0	100.0	94.0	100.0
	%se				100.0	92.2	100.0	100.0	94.0	100.0
Rsp=0.98	%sp	100.0	94.0	100.0	100.0	76.2	100.0			

Table B. Visit 2.

Principle 3		20.0			55.0			80.0		
		90% Conf.interval			90% Conf.interval			90% Conf.interval		
		est.	lower	upper	est.	lower	upper	est.	lower	upper
	%equal	100.0	94.1	100.0	100.0	94.1	100.0	100.0	94.1	100.0
	%se				100.0	94.1	100.0	100.0	94.1	100.0
Rsp=1.00	%sp	100.0	94.1	100.0						

Principle 2		20.0			55.0			80.0		
		90% Conf.interval			90% Conf.interval			90% Conf.interval		
		est.	lower	upper	est.	lower	upper	est.	lower	upper
	%equal	98.0	90.7	99.9	98.0	93.7	90.7	100.0	94.1	100.0
	%se	75.0	24.9	98.7	100.0	93.1	100.0	100.0	94.1	100.0
Rsp=0.98	%sp	100.0	93.6	100.0	85.7	47.9	99.3			

Table 3.1.6.

Comparison of criterion scores using the full model or simplification strategy model 1 (predicting gait scores from measures of hock burn on-farm), indicated as distribution of flocks across final flock scores as defined by Welfare Quality® including estimates for agreement between methods (Est), and the lower (Low) and upper (Upp) confidence limits within classification groups for equality (% eq);specificity (%sp), sensitivity (%se), Table A shows the result for visit 1, table B shows the results for visit 2.

Table A. Visit 1.

Criterion 6		20.0			55.0			80.0		
		90% Conf.. interval			90% Conf.. interval			90% Conf. interval		
		est.	lower	upper	est.	lower	upper	est.	lower	upper
	%equal	95.7	87.2	99.2	100.0	93.8	100.0	100.0	93.8	100.0
	%se	97.3	87.8	99.9	100.0	93.8	100.0	100.0	93.8	100.0
Rsp=0.98	%sp	90.0	60.6	99.5						

Table B. Visit 2.

Criterion 6		20.0			55.0			80.0		
		90% Conf.. interval			90% Conf.. interval			90% Conf. interval		
		est.	lower	upper	est.	lower	upper	est.	lower	upper
Morgen	%equal	93.9	84.9	98.3	100.0	94.1	100.0	100.0	94.1	100.0
	%se	94.6	94.0	99.0	100.0	94.1	100.0	100.0	94.1	100.0
Rsp=0.97	%sp	91.7	66.1	99.6						

Table 3.1.7.

Comparison of criterion scores using the full model or simplification strategy model 2 (predicting clinical scores and gait score on-farm from slaughterhouse measures of footpad dermatitis and hock burn), indicated as distribution of flocks across final flock scores as defined by Welfare Quality® including estimates for agreement between methods (Est), and the lower (Low) and upper (Upp) confidence limits within classification groups for equality (% eq);specificity (%sp), sensitivity (%se), Table A shows the result for visit 1, table B shows the results for visit 2.

Table A. Visit 1.

Criterion 6		20.0			55.0			80.0		
		90% Conf. interval			90% Conf. interval			90% Conf. interval		
		est.	lower	upper	est.	lower	upper	est.	lower	upper
	%gelijk	100.0	93.8	100.0	100.0	93.8	100.0	100.0	93.8	100.0
	%se	100.0	92.2	100.0	100.0	93.8	100.0	100.0	93.8	100.0
Rsp=1.00	%sp	100.0	74.1	100.0						

Criterion 3		20.0			55.0			80.0		
		90% Conf. interval			90% Conf. interval			90% Conf. interval		
		est.	lower	upper	est.	lower	upper	est.	lower	upper
	%equal	93.8	84.6	98.3	81.2	69.6	89.8	100.0	94.0	100.0
	%se	33.3	1.7	86.5	96.7	85.1	99.8	100.0	94.0	100.0
Rsp=0.92	%sp	97.8	89.9	99.9	55.6	34.1	75.6			

Table B. Visit 2.

Criterion 6		20.0			55.0			80.0		
		90% Conf. interval			90% Conf. interval			90% Conf. interval		
		est.	lower	upper	est.	lower	upper	est.	lower	upper
	%equal	100.0	94.1	100.0	100.0	94.1	100.0	100.0	94.1	100.0
	%se	100.0	92.2	100.0	100.0	94.1	100.0	100.0	94.1	100.0
Rsp=1.00	%sp	100.0	77.9	100.0						

Criterion 3		20.0			55.0			80.0		
		90% Conf. interval			90% Conf. interval			90% Conf. interval		
		est.	lower	upper	est.	lower	upper	est.	lower	upper
	%equal	98.0	90.7	99.9	83.7	72.5	91.6	100.0	94.1	100.0
	%se	85.7	47.9	99.3	100.0	92.2	100.0	100.0	94.1	100.0
Rsp=0.92	%sp	100.0	93.1	100.0	38.5	16.6	64.5			

To explore the potential of nature to improve the quality of life

Wageningen UR Livestock Research
P.O. Box 338
6700 AH Wageningen
The Netherlands
T +31 (0)317 480 10 77
E info.livestockresearch@wur.nl
www.wageningenUR.nl/livestockresearch

Livestock Research Report 833

Together with our clients, we integrate scientific know-how and practical experience to develop livestock concepts for the 21st century. With our expertise on innovative livestock systems, nutrition, welfare, genetics and environmental impact of livestock farming and our state-of-the art research facilities, such as Dairy Campus and Swine Innovation Centre Sterksel, we support our customers to find solutions for current and future challenges.

The mission of Wageningen UR (University & Research centre) is 'To explore the potential of nature to improve the quality of life'. Within Wageningen UR, nine specialised research institutes of the DLO Foundation have joined forces with Wageningen University to help answer the most important questions in the domain of healthy food and living environment. With approximately 30 locations, 6,000 members of staff and 9,000 students, Wageningen UR is one of the leading organisations in its domain worldwide. The integral approach to problems and the cooperation between the various disciplines are at the heart of the unique Wageningen Approach.