Development of the 'animal welfare' dimension within the Greenwell sustainability assessment model: 2. Data for all production stages and calculation of scores for the broiler on-farm stage

I.C. de Jong, D. te Beest

Report 1260

WAGENINGEN
UNIVERSITY & RESEARCH

Development of the 'animal welfare' dimension within the Greenwell sustainability assessment model: 2. Data for all production stages and calculation of scores for the broiler on-farm stage

Ingrid de Jong[1] and Dennis te Beest[2]

1 Wageningen Livestock Research, Wageningen University and Research
2 Biometris, Wageningen University and Research

This research was conducted by Wageningen Livestock Research, commissioned and funded by the Ministry of Agriculture, Nature and Food Quality, Aviagen EPI, Belgabroed/Van Hulst, De Heus BV, and Plukon Food Group within the framework of the public-private partnership 'Greenwell' (TKI-AF-17023, project number BO-47-001-033).

Wageningen Livestock Research
Wageningen, June 2020

Report 1260

WAGENINGEN
UNIVERSITY & RESEARCH

De Jong, I.C., te Beest, D. 2020. *Development of the Greenwell welfare assessment model: 2. Data for all production stages and calculation of scores for the broiler on-farm stage.* Wageningen Livestock Research, Report 1260.

Samenvatting Binnen het Greenwell project wordt een model ontwikkeld om de effecten van vleeskuikenproductiesystemen op duurzaamheid (milieu, economie, dierenwelzijn en diergezondheid) te kunnen inschatten. In dit rapport wordt een overzicht gegeven van data verzameld in de productieketen voor de diverse welzijnsindicatoren in de verschillende productiefases van vleeskuikens (vleeskuikenouderdieren, broederij, vleeskuikenbedrijf, vangen, transport en slacht) op basis van het Greenwell welzijnsmodel voor vier geselecteerde vleeskuikenproductiesystemen (conventioneel, Kip van Morgen, Beter Leven 1 ster, biologisch), indien beschikbaar. Deze data zijn verzameld of er is een inschatting gemaakt door de Greenwell consortiumpartners of andere ketenpartners in Nederland voor het jaar 2017, met soms aanvullende informatie uit 2018 en 2019. In aanvulling daarop is een scoresysteem ontwikkeld om een welzijnsscore op een schaal van 0 (slecht) tot 100 (perfect) toe te wijzen voor de diverse indicatoren voor het welzijn van vleeskuikens op het primaire bedrijf. Daartoe is gebruik gemaakt van een dataset uit 2017 en 2018 van een slachterij aangevuld met data afkomstig van een broederij. De som van deze scores geeft de totale welzijnsscore voor een koppel vleeskuikens op het primaire bedrijf. Resultaten laten zowel de gemiddeldes per productiesysteem zien, als de variatie binnen een productiesysteem, voor conventioneel, Kip van Morgen, en Beter Leven 1 ster. Er waren geen data beschikbaar voor biologische koppels en derhalve kon voor dit productiesysteem geen welzijnsscore worden bepaald.

Summary UK. Within the Greenwell project, a model will be developed to assess the sustainability (environmental impact, economic impact, animal welfare and health impact) of broiler production systems. The present report provides values and variation for the welfare indicators in the different stages of the broiler production chain (broiler breeders, hatchery, broiler farm and end-of-life stage) based on the Greenwell welfare model for four selected production systems (conventional, Dutch Retail Broiler, Better Life one star, organic), if known. These data have been collected by the Greenwell consortium partners or other chain partners in the Netherlands over the year 2017; where no data were available estimations based on 2018 or 2019 were used. In addition to that, a scoring system has been developed to assign a welfare score on a scale between 0 (worst)-100 (perfect) for the welfare indicators of broiler production on-farm. To this end, a database containing processing plant and hatchery data over 2017 and 2018 has been used. The sum of the scores for the individual indicators presents the total welfare score for a particular broiler flock. Results show the average welfare scores for broilers on-farm and the variation within production systems, for conventional, Dutch Retail Broiler and Better Life one star. No data were available to assign scores to organic flocks.

This report can be downloaded for free at https://doi.org/10.18174/524330 or at www.wur.nl/livestock-research (under Wageningen Livestock Research publications).

Table of contents

Summary

Within the Greenwell project, a model will be developed to assess the overall sustainability (environmental impact, economic impact, animal welfare and health impact) of broiler production systems. The present report provides values and variation for the welfare indicators in the different stages of the broiler production chain (broiler breeders, hatchery, broilers on-farm and end-of-life stage) based on the Greenwell welfare model for four selected production systems (conventional, Dutch Retail Broiler, Better Life one star, organic) as published previously. Data collected by the project partners over the years 2017 and 2018 have been used; where no data was available others were consulted or expert opinion was used to provide an estimation. Most data were collected for broilers on-farm, followed by the broiler breeder phase, whereas for the end-of-life phase and hatchery phase the collected data were limited to very limited, respectively. The values and variation for the different indicators resulted in a general overview of the welfare performance of the four selected production systems and suggest a favourable welfare for broiler breeders and broilers on-farm for Dutch Retail Broiler and Better Life one star as compared to conventional. For organic insufficient data could be collected at all to draw any conclusions about welfare.

The present report also describes the development of a welfare scoring system for broilers on-farm. The development of the Greenwell welfare assessment model was explained in a previous report; for broiler chickens on-farm initially 11 welfare indicators were selected of which seven were animal based (footpad dermatitis, hock burn, breast irritation, liveability, scratches, locomotion and % broilers resting on an elevated structure) and four were resource or management based indicators (maximum stocking density, provision of early feeding and water, antibiotics usage and an overall score for presence of environmental enrichment, natural light, covered veranda or outdoor range). To be able to compare the individual welfare indicators and to calculate an overall welfare score, calculations were developed to score these various welfare indicators of the broiler farm stage on a similar scale, i.e., from 0 (lowest welfare level) to 100 (highest welfare level). This was done for eight welfare indicators of the original model (five animal-based indicators, i.e., footpad dermatitis, hock burn, breast irritation, liveability, scratches, and three resource-based indicators, i.e., maximum stocking density, provision of early feeding and water, and an overall score for presence of environmental enrichment, natural light, covered veranda or outdoor range). 'Locomotion' could not be included because data were not collected on a routine basis. 'Curative antibiotics use' was excluded from the original model as data showed a relation with liveability. The 'proportion of broilers resting on an elevated structure' was excluded because this data are not collected routinely, and including a score for the presence of an elevated resting structure would result in a double counting as this is already included in the enrichment score.

The Welfare Quality® method was used to assign scores to each welfare indicator on a scale from 0-100. Flock data of the processing plant and the hatchery, collected in 2017 and 2018 for 5683 conventional, 5936 Dutch Retail Broiler and 1889 Better Life one star broiler flocks, were used to develop the calculations in case these were not included in Welfare Quality®, or when Welfare Quality® could not be applied because of differences in scoring methodology. Finally, a model to calculate the total welfare score for broilers on-farm was developed resulting in median welfare scores of 374, 460 and 602 for conventional, Dutch Retail Broiler and Better Life one star respectively (scores on a scale from 0 (worst) to 800 (best)). When the total welfare score is based on the 5 animal-based welfare indicators only, the median welfare score was 331, 370 and 396 for conventional, Dutch Retail Broiler and Better Life one star respectively (on a scale from 0 (worst) to 500 (best)). Data for organic production were lacking and scores for this production system could therefore not be calculated.

Flock data were also available for two indicators of broiler welfare during the end-of-life stage. Because data of most indicators for this production stage were lacking, calculations for only these two indicators (Dead-on-Arrival, proportion of bruises) were developed on a scale between 0 to 100, but no calculation for the total welfare score during the end-of-life stage was developed. This remains to

be done, when more data will become available. Because of lack of routinely collected data, no welfare scoring system could be developed for the broiler breeder and hatchery stage.

The Greenwell welfare model for broiler welfare on-farm showed, based on data of 2017 and 2018, a better welfare score of flocks in Better Life one star over Dutch Retail Broiler and conventional production systems respectively. It also illustrated that scores within production systems show a high variability and that scores between production systems showed considerable overlap, especially when the resource- and management based indicators were removed from the model. This variation in welfare scores suggests room for improvement of broiler welfare within production systems and provides options to use this welfare assessment model to assist improvement of broiler welfare on individual farms by e.g. management advice. It should be noted that locomotion data (gait scores) could not be included because these data are not collected routinely. As locomotion problems is one of the most important welfare issues in broiler chickens, collection of these data and inclusion in the welfare score for each production system is highly recommended.

The Greenwell welfare model should not be considered as a static model. It can be improved as soon as additional data become available, e.g. data on the actual behaviour of the birds. An equal weighing to all indicators is applied because there is currently insufficient knowledge on the exact effects of the individual indicators on broiler welfare, but this may be subject to change in the future. In addition, simply summing the scores allows compensation between indicators which may not be preferred, and this can be adjusted in the future by e.g. defining minimum values for certain indicators.

To conclude, the Greenwell welfare model for broilers on-farm takes into account the most important welfare issues and provides a transparent way to calculate a total welfare score. The generated welfare score for broiler flocks can be used to compare systems and flocks with respect to their welfare performance. It can be considered as a starting point for assessing welfare on a large scale as part of an overall sustainability assessment of broiler production.

1 Introduction

One of the aims of the Greenwell project [1] is to develop a model to assess the sustainability of broiler production systems. Production systems being defined as standardised housing and management procedures or requirements, including the use of genetic strains with specific qualities or limitations on average daily weight gain for strains that are allowed. Such a sustainability model can be used to provide insight in differences between production systems with respect to sustainability, to substantiate choices for production systems, and to determine the effect of system adaptations (e.g., in resources, housing conditions) and their respective impact on overall sustainability. The ultimate goal of the Greenwell project is to develop a model that can be used for real-time data collection on sustainability aspects of broiler production systems [1]. Therefore, the model should not only provide insight in differences between production systems with respect to sustainability aspects, but also provide insight in variation within broiler production systems (between flocks and across time).

Within the Greenwell project, the choice was made to compare the wide range of broiler production systems that are currently present in The Netherlands [2, 3], and data will be shown for these production systems. Four systems were selected that are supposed to represent the range between 'efficient in terms of production and costs', and 'including additional requirements supposed to provide a higher welfare level for the chickens' [2, 4]:
(1) the conventional broiler production system using so-called fast growing breeds housed at maximum stocking densities (in the Netherlands: 39-42 kg/m^2) and with indoor housing only ('**conventional**'), representing the majority of broiler chicken production in the Netherlands;
(2) systems according to the production standards of 'Kip van Morgen' [5], i.e. a slower growing chicken breed with a maximum daily growth rate of 50 g and a stocking density of 38 kg/m^2 or lower, and provision of environmental enrichment in the house but no veranda or outdoor range, called '**Dutch Retail Broiler**' (DRB);
(3) **Better Life one star** (BLS) ('Beter Leven' label of the Dutch Animal Protection Organisation), using a slower growing breed (slaughter age at least 56 days), a stocking density of max 25 kg/m^2, a covered veranda and environmental enrichment, sometimes also referred to as 'free range indoor' and
(4) **organic**, using a slow growing breed (slaughter age of at least 70 days), a stocking density of max 21 kg/m^2 and an outdoor range.

The development of the welfare assessment model, i.e. the selection of indicators for the broiler breeder, hatchery, broiler on-farm and end-of-life stage, within the Greenwell sustainability model has been reported in [6]. The aim of the current report is two-fold:

(1) Firstly, it will show the **values and variation for the different indicators in the different stages of the production chain for the four selected production systems**, if available. These data have been collected or estimated by the consortium partners or other chain partners in the Netherlands, and we have tried to collected as many data as possible over the year 2017. In some cases, data or estimations (expert opinion) based on other years (2018 and 2019) had to be used. These data on the level of the indicators will provide a general overview of the welfare performance of the four selected production systems.
(2) Secondly, we will **apply and develop calculations to score the various welfare indicators of the broiler on-farm stage** on a similar scale, i.e., from 0 (lowest welfare level) to 100 (highest welfare level), so that these can be compared for the different indicators. To this end, we will use the Welfare Quality method [7] to assign scores to each welfare indicator, according to the welfare model for broilers on-farm as described in [6]. Flock data of the processing plant and hatchery, collected in 2017 and 2018, will be used to develop calculations in case these are not included in Welfare Quality, or when Welfare Quality cannot be applied because of differences in scoring methodology. Finally, a model to calculate the total welfare score for broilers on-farm will be developed and average scores will be presented for three production systems (conventional, Dutch Retail Broiler and Better Life one

star) for the years 2017 and 2018. Data for organic production were lacking and scores for this production system could therefore not be calculated.

Flock data are also available for two indicators of broiler welfare during the end-of-life stage. Because data of most indicators for this production stage are lacking, we will only develop calculations for these two indicators on a scale between 0 to 100, but we do not develop a calculation for the total welfare score during the end-of-life stage. This remains to be done, when more data will become available. Because for the broiler breeder and hatchery stage many data are lacking, as these are currently not collected on a routine basis, we do not develop scoring systems yet to assess welfare during these stages of the broiler production chain.

2 Materials and methods

2.1 Part 1: data collection for the four selected production systems

2.1.1 Data collection methods

For each production stage (broiler breeders, hatchery and day-old chick transport, broilers on-farm and broilers end-of-life), data were collected for the four selected production systems (conventional, Dutch Retail Broiler, Better Life one star, organic) over the year 2017. Data were provided by the Greenwell project partners that together represented all production chain phases and for the broiler breeder stage also from an external source (breeder processing plant). If no data of 2017 were available this is indicated in the respective tables for the production stages. For some data of the broiler breeder stage, expert opinion was used to make an estimation. This expert opinion was based on the current (2018/2019) situation whereas the data were available from 2017. Below we describe how data were collected and the specific data sources, and in the tables in the results section the data source and year of collection are also indicated.

It should be noted that data could not only comprise Dutch farms, but also farms in surrounding countries being part of the production chain (e.g., broiler farms delivering chickens to processing plants located in The Netherlands). These farms did meet the criteria of the four defined production systems. It is also important to keep in mind that the data that are presented for the breeder stage do not necessarily represent the parents of the chickens of which data in the hatchery, broiler on-farm and end-of-life stage were collected, as for the general tables in section 3.1. we aimed to collect as much information as was possible for all production chain phases. For part 1 of the report, no total welfare score has been developed, neither per chain phase. For this we refer to part 2 of the report.

As most data are available for the broiler on-farm stage, in the following paragraphs and sections we will always start with this stage followed by the broiler breeder, hatchery and end-of-life stage.

2.1.1.1 Data collected for broilers (on-farm)

Data were collected at flock level. A broiler flock was defined as all birds originating from the same house at the broiler farm, being placed at one date and depopulated at the same time. For each indicator it is described how the data were collected. Data can be collected at thinning (for conventional systems) and the end of the production cycle, we here only used data that have been collected upon the end of cycle.

Provision of early feeding and water

These measures served as an alternative to the proportion of emaciated chickens, as data on emaciation are not collected on a routine basis and thus were not available. Data were provided by the hatchery for on-farm hatching, as they register whether or not 18-days incubated eggs or day-old chickens will be delivered to the farm. The proportion of flocks receiving early feeding in the hatchery has been estimated by the processing plant (no registration available). All data were collected or estimated over the year 2017.

Proportion of broilers resting on an elevated structure

These data are not being collected on a routine basis and there is no other source of information available representing the commercial situation.

Footpad dermatitis

Footpad dermatitis data were collected by the processing plant for each flock at depopulation, according to national legislation. In short, per flock a sample of 100 right feet was collected (50 feet at

approximately 1/3 of the flock and 50 feet at approximately 2/3 of the flock) and scored in three classes: score 0, no lesion (no or very small discoloration of the feet); score 1, mild lesion (mild lesion, superficial discoloration of the skin, and hyperkeratosis); score 2, severe lesion (epidermis affected, blood scabs, haemorrhage, and severe swelling of the skin) [8]. These data were collected in 2017. For organic flocks, footpad lesions have not been collected on a routine basis. To get an impression of footpad dermatitis in organic flocks, 10 randomly chosen flocks were sampled in February 2019 and 10 additional flocks in August 2019 according to the method as described above.

Hock burn
Hock burn data were collected by the processing plant for each flock at depopulation, according to the national quality assurance scheme for processing plants [9]. Per lorry, 100 broilers should be scored for the presence of hock burn by inspecting both hocks. Broilers receive a score of 1 (presence of hock burn) when a dark coloured area of >0.5 cm^2 on the hock is present. No data were routinely collected for organic flocks and thus were not available.

Breast irritation
Breast irritation data were collected by the processing plant for each flock at depopulation, according to the national quality assurance scheme for processing plants [9]. Per lorry, 100 broiler carcasses should be scored for the presence of a brown/black area on the breast of 2 cm^2 or larger, which receive a score of 1 (breast irritation present). No data were routinely collected for organic flocks and thus were not available.

Maximum stocking density
For maximum stocking density, the requirements of the production systems ('Dutch Retail Broiler' and 'Better Life one star') [3] or national/European legislations were used (conventional (Directive 2007/43/EC) and organic (Council Regulation (EEC) No 2092/91)), as there is no database containing the information of producers on the actual stocking densities that have been applied.

Locomotion
Quality of locomotion was assessed using the gait score [7]. Briefly, broiler chickens between 1-5 days before depopulation were assessed for locomotor quality using a score from 0 to 5, a score of 0 meaning 'perfect' and a score of 5 'unable to walk'. No routinely collected data were available. As the data available from trials under commercial conditions were relatively old (most data originating from 2013-2014) these were not considered representative for the situation in 2017 due to genetic selection for locomotion. Therefore, data collected by a project partner in 2019 have been used. These data were collected under experimental conditions (pen sizes larger than 25 m^2) or semi-commercial conditions (7 flocks of 12.000 chickens during 2019).

Mortality
Mortality data were collected by the farmers and registered on the food chain information form ('VKI formulier') when birds were sent to processing at depopulation. The mortality figure of a flock comprised all birds found dead and birds culled, and includes early (first week) as well as late mortality. For organic flocks, no registration on flock base was available, but the processing plant provided the estimated average mortality figure.

Curative antibiotics use
Since the start of the National monitoring program for antimicrobial treatments in various species, the application of preventive microbial treatments has been phased out [10]. We therefore considered all treatments to be curative. Farmers register the use of antimicrobials on the food chain information form, and the processing plant registered whether or not antimicrobials were used for a specific flock. No data were available for organic flocks.

Scratches
Data on scratches and wounds were collected by the processing plant for each flock at depopulation, according to the national quality assurance scheme for processing plants [9]. Per lorry, 100 carcasses should be inspected (back or thigh area). A score of 1 was assigned when 3 scratches > 2cm were

observed (fresh, or scab or crust) or when a wound (open skin, either or not covered with a crust) was observed. No data were available for organic flocks.

Environmental enrichment, natural light, covered veranda and outdoor range
Presence of environmental enrichment, natural light, covered veranda and outdoor range served as an alternative to the registration of the proportion of time spent on species specific behaviour. The proportion of flocks with no, 1, 2 or 3 enrichments, covered veranda, outdoor range and natural light was provided by the plant. Environmental enrichments comprised bales, pecking stones, scattering grains in the litter, and elevated resting areas (perches, platforms). These data are registered as part of specific production requirements for part of conventional flocks or all ('Dutch Retail Broiler' [11], Better Life one star and organic flocks.
(https://beterleven.dierenbescherming.nl/fileupload/2019/201904_Factsheet_BLk_vleeskuikens_update_DEF.pdf)) production systems.

2.1.1.2 Broiler breeders

Breed/line
The breeding and nutrition company provided information on whether or not fast growing males/females or slower-growing/dwarf females were used within the different broiler production systems.

Water provision time
For the rearing and production period, the proportion of farms applying unrestricted or restricted water supply was provided by the hatchery and/or estimated by the nutrition company.

Proportion of breeders resting on an elevated structure
No data have been collected or available from literature for both rearing and production.

Footpad dermatitis and breast blisters
Footpad dermatitis and breast blister scores were collected by one Dutch processing plant, processing the majority of the broiler breeders housed in The Netherlands. The data included conventional parent stock flocks and parent stock of 'Dutch Retail Broiler' and 'Better Life one star'. As only the breed was registered by the plant, the link to the system has been provided by other chain partners using the date of birth of the flock and the production farm identification. Data were anonymised before further processing. Both footpad dermatitis and breast blisters were scored on a range from 1-5, a score of 1 being no abnormalities observed until a score 5 being many abnormalities observed. It should be noted that this was a qualitative score by processing plant personnel and that no counts were performed. Further, they were not allowed to score the middle category (score 3). In case a flock was processed in different batches, the average score of these batches was provided.

Maximum stocking density
Stocking densities generally applied during rearing and production, presented as females and males per m^2, were provided by the nutrition company.

Locomotion
No data have been collected or available from literature for both rearing and production.

Total mortality
Estimated figures of total mortality per type of production system in rearing and production were provided by the nutrition company.

Curative antimicrobials use
For the production period, these data were provided by the hatchery and comprised the number of flocks with or without antibiotic treatments in the production period. In addition, results of the national monitoring of antibiotics usage were used [12].

Injuries
No data have been collected or available from literature for both rearing and production.

Proportion of flocks without mutilations
Estimates per production system were provided by the nutrition company.

Environmental enrichment, covered veranda, outdoor range and natural light
These data were collected as an alternative indicator for the proportion of time spent on species-specific behaviour. For the production period, these data were provided by the hatchery by an inventory of the production farms that delivered eggs to the hatchery. Thus, data were collected on a farm base. Environmental enrichment comprised bales of lucerne, pecking stones, distribution of grains in the litter; each was counted as one enrichment.

2.1.1.3 Hatchery

Only few data were routinely collected for the hatchery stage. The proportion of flocks receiving feed and water in the hatchery was estimated by the processing plant (see 2.1.3.1.). The proportion of flocks hatched on-farm served as alternative measure for the proportion of chickens showing undisturbed resting, hatchery procedures (culling procedures, sexing, vaccinations, disinfection at the hatchery) and the proportion of chickens showing species specific behaviour. These data were provided on a flock base by the hatchery.

2.1.1.4 Broilers – end of life stage

Body weight loss, water withdrawal time and stocking density in containers
These data are not recorded on a routine basis. Feed withdrawal time has been suggested as an alternative to body weight loss, but these data are also not collected on a routine basis. Stocking density in containers is suggested as alternative indicator for behaviour in containers. There were no other data sources, that included a sufficient number of flocks of the different production systems, which could be used.

Dead-on-arrival
These data are recorded for each flock upon arrival at the processing plant, according to the routine ante-mortem inspection of each flock before processing.

Proportion of chickens with trapped limbs, supine broilers, splayed legs and wing/leg fractures
These data are not recorded on a routine basis and there were no other data sources, that included a sufficient number of flocks of the different production systems, which could be used.

Proportion of broiler chickens with bruises
Bruises are scored according to the national quality control guidelines for processing plants [9]. Carcasses are inspected at two time points (at 1/3 and 2/3 within the flock) during two minutes each time, from the breast side of the carcass. All bruises with a dark red colour and a size larger than 3 cm^2 on the breast, wings and legs are scored as 'bruise present'.

Consciousness when shackling and stunning method
These data are dependent on the respective processing plant and the applied stunning and killing method, and are provided by the plant.

2.2 Part 2: calculation of welfare scores for indicators of the on-farm broiler stage

2.2.1 Data

To develop calculations to score the different indicators and the total welfare score of a flock, the processing plant provided data that were collected on a routine basis in 2017 and 2018 for three

production systems (conventional, Dutch Retail Broiler, Better Life one star). The database comprised the following indicators: footpad dermatitis (%), hock burn (%), scratches (%), mortality (%), breast irritation (%), Dead-on-Arrival (DoA) (%), bruises on the wings, legs and breast (%). These data were collected as described in section 2.1 of this report. In addition, the processing plant and the hatchery provided data of the application of on-farm hatching or early feeding in the hatchery. For stocking density and environmental enrichment/natural light/covered veranda or outdoor range, we used the requirements of the various labelling programs or legislation and assigned a standard value to a particular production system.

2.2.2 Development of calculations for indicator scores

For calculation of the scores for indicators for broiler welfare on-farm, the Welfare Quality method [7] was used as a starting point. Welfare Quality developed methods to calculate scores for each indicator, and for the 12 welfare criteria and four welfare principles. These latter two (criterion and principle score calculations) will not be applied here. We will only apply the calculation of the indicator score, in case data have been collected in a similar way as defined by Welfare Quality [7].

For each indicator, scores will be calculated on a scale between 0 and 100, with 0 representing the worst situation and 100 the best. This enables us to compare scores for different indicators. Table 1 shows the calculations according to Welfare Quality that have been applied in the current study [7, 13]. Appendix 1 provides the graphs showing the relationship between the index and the score.

Table 1 *Calculation of scores for indicators in case the Welfare Quality protocol could be applied [7, 13]. See also the graphic presentations in Appendix 1.*

Indicator	Index calculation	Score calculation	Reference
Stocking density	I = ((100/(42-4) x (42-d) [1]	When I≤ 30: Score = (2.6077 x I) – (0.051672 x I^2) + (0.00050863 x I^3) When I ≥ 30: Score = (12.019) + (1.4058 x I) – (0.011609 x I^2) + (0.000063483 x I^3)	[7]
Footpad dermatitis	Index=(100-(2*%score 1)+ (7*%score 2))/7[2]	When I ≤ 70: Score = (0.50686 x I_{fpd}) - (0.0072409 x I_{fpd}^2) + (0.000081315 x I_{fpd}^3) When I ≥ 70: Score = - 513.33 + (22.507 x I_{fpd}) - (0.32152 x I_{fpd}^2) + (0.0015779 x I_{fpd}^3)	[7]
Locomotion (gait score)	Index= 100 – (% moderately lame birds / 5) - % severely lame birds[3]	When I ≤ 80 then Score = (0.28221 x I_{lame}) - (0.0029368 x I_{lame}^2) + (0.000041416 x I_{lame}^3) When I ≥ 80 then Score = -3822.8 + (143.64 x I_{lame}) - (1.7949 x I_{lame}^2) + (0.0075078 x I_{lame}^3)	[7]
Liveability	Index = 100 - % mortality	When I≤85: Score = (0.024147 x I) + (0.0003195893 x I^2) - (0.0000033715 x I^3) When I≥85 and I≤95: Score = -21268.82 + (750.6883 x I) – (8.83102368 x I^2) + (0.0346293473 x I^3) When I≥95: Score = 109906.77 - (3391.6987 x I) + (34.77305 x I^2) - (0.1183674060 x I^3)	[13]

[1] d=maximum stocking density in kg/m^2 ; 42 and 4 represent the maximum and minimum stocking density respectively in kg/m^2. In the original calculation a maximum stocking density of 44 kg/m^2 was included, this has been adjusted according to current European legislation, based on the advice of the Welfare Quality broiler working group (de Jong, pers. comm.).

[2] Welfare Quality originally scores footpad dermatitis into 5 classes. %Score 1 represent mild footpad dermatitis (Welfare Quality scores 1+2; equal to processing plant score 1) and %Score 2 represent severe footpad dermatitis (Welfare Quality scores 3+4; equal to processing plant score 2).

[2] Moderately lame: gait score 3; severely lame: gait score 4 and 5.

In case the assessment method of an indicator differed as compared to Welfare Quality, or the indicator has not been included in the Welfare Quality broiler assessment protocol, new score calculation methods needed to be developed. To develop these new score calculations, we applied the same methodology as Welfare Quality, i.e., international experts were consulted and asked to provide scores for farms with different values for the various indicators, and subsequently spline functions

were calculated based on the expert opinions (see chapter 7 in [14] for a description of the methodology).

Per variable we asked experts to evaluate a set of 13-14 data points. The evaluated points were handpicked based on the distribution of the respective variable. Most of the evaluated points were located in the area where the variable has its highest mass. At the same time we also wanted to cover the observed range without making the distance between two consecutive evaluation points too large. Per variable, we calculated the average across the experts per evaluated point. We then fitted a spline through these averaged values. There are several criteria the spline fit should meet: (1) If the variable has its best score, the spline should return a score of 100, (2) The spline should be monotonously decreasing as the variable of interest is further away from it best value, and (3) the spline function should be strictly positive (e.g. it should not return negative values). We used R package *cobs* for fitting of the splines. The advantage of this package is that we can explicitly specify the mentioned three fitting criteria. Appendix 2 shows the questionnaire as has been sent to the experts; Appendix 3 includes information on the experts that have been consulted; Appendix 4 shows the graphical representation of the scores as provided by the different experts, and the spline function that was calculated according to these scores. Scores for early feeding and water, and for enrichment/natural light/covered veranda or outdoor were determined by applying a decision tree. These scores are included in Appendix 4 as well.

In addition to the current report, a software tool in R has been developed that can be used to calculate scores for the various indicators and a total flock welfare score. The new splines are therefore not represented as a function here but only as a graphical representation in Annex 4.

Two adjustments were made in comparison to the welfare model developed in de Jong [6]. Because elevated platforms are included in the total enrichment score, a score for broilers resting on an elevated area has not been developed, as this could result in a double-counting. In case in the future data on the actual use of enrichment and behaviour can be collected, the resource-based decision trees can be replaced by animal-based measures and resting on an elevated area can then be separately scored.

Secondly, antibiotics usage has been proposed to be included in the welfare model [6]. This is a management-based measure rather than an animal-based measure. Antibiotics usage has been classified as yes or no but not as a (semi)continuous variable based on the average daily dosage, which can have a relatively large effect on the total score for a flock. Furthermore, inspection of the combined histograms of antibiotics usage and mortality indicated that a different antibiotics application strategy seemed to be followed for the different broiler production systems, indicating at least a different threshold in mortality before antibiotic treatments are applied. The histograms are shown in Appendix 5 and may suggest that in conventional systems antibiotics is already applied with low mortality figures, whereas in BLS antibiotics is not applied until a relatively high mortality is observed. It was therefore further explored whether mortality and antibiotics application could be combined into one score for 'absence of disease'. To that end, four different approaches were explored: (1) no antibiotics applied received a score of 100, antibiotics applied received a score of 0, and the average of the antibiotics and mortality score was used; (2) similar approach, but no antibiotics received a score of 80 points and antibiotics a score of 30 points; (3) <3% mortality and no antibiotics received 80 points, <3% and antibiotics received 20 points, >3% mortality 0 points and the average between the mortality and antibiotics score is used; (4) similar approach, but <3% and antibiotics also received 0 points and again the average between the mortality and antibiotics score is used. Options 1, 2 and 4 resulted in suboptimal distribution of flocks over the score between 0-100; only option 3 resulted in a relatively better distribution of flocks over the scores between 0-100. However, in option 3 mortality is also double counted which results in mortality having a relatively large effect on this score. Considering this, and the fact that it is a management-based measure for health and health is also reflected in the total mortality, it was decided not to include antibiotics usage in the final model.

Finally, gait score data to assess locomotion are included in the final model but because these data are not being collected routinely, we could not calculate these scores for the same flocks as for the other

data. Therefore, we will present the calculations for gait score as an example using data collected by a project partner in an experimental setting, based on Table 1, but do not include these data in the total flock score.

According to the existing or new spline functions, scores have been calculated per indicator for each production system using the data as collected by the processing plant and presented in tables with the minimum, maximum, 10%, 50% and 90% percentile values for the various production systems.

2.2.3 Development of the total welfare score for broilers on-farm

As the assessment protocol for broiler welfare on-farm includes a number of indicators, we first determined whether or not we could reduce the number of indicators included by calculating the correlation coefficient between all indicators that are routinely collected. In case a high correlation will be found between two indicators, the value of one indicator could be predicted by the correlated indicator according to [13]. The correlations between variables were calculated using Spearman's correlation coefficient. In visualizing the correlation between two variables (e.g. the scatter plots) we first transformed the data by taking the square root. The advantage of this transformation is that lower values are emphasized more in the figure and the effect of a long right tail is reduced. Because we did not find a high correlation between indicators, simplification of the broiler on-farm assessment model was not possible (see results section for correlation coefficients). We thus continued with all indicators according to the developed model [6] for which we had received data.

The overall welfare score per farm was determined by simply summing up all scores for the individual indicators, which were eight indicators because locomotion (gait score) data were not collected on a routine basis. The maximum flock score that could be assigned is then determined by the indicators for which a score can be calculated, i.e., 800 points. In case gait score data are also collected, the maximum score will be 900 points. Two reasons were taken into account to assign an equal weight to each indicator for the total flock score. Firstly, as simply summing up all individual scores resulted in substantial variation in scores within and between flocks for the three production systems, we assumed that we had a sufficiently sensitive method to discriminate between the different production systems. Secondly, there are, based on the scientific literature, insufficient arguments to assign a different weighing to each indicator based on the effect of the individual indicator on broiler welfare[1]. A further sensitivity analysis of the total welfare score was performed by omitting the scores for resource- and management based measures (stocking density, early feeding, enrichment) one-by-one or in pairs and comparing these with the total farm score including all measures.
Finally, for each indicator, and the sum of all indicators of which data were available, we calculated the score per flock. Per production system, the minimum, maximum, 10%, 50% and 90% percentiles were calculated and are presented in the results section, in addition to histograms presenting the distributions within and between production systems. Differences in median value between production systems were tested with Mood's median test. Mood's median test is non-parametric and tests whether the medians of the two systems are identical. In addition, the Wilcoxon rank-sum test (also called Mann–Whitney U test) was done on the scores. This test compares the sum of ranks of two systems and tests whether the two systems have a similar distribution.

2.2.4 Total welfare score for the end-of-life stage

For the end-of-life stage, we only had data of two indicators, i.e., the proportion of birds Dead-on-Arrival and the proportion of chickens with bruises (on the legs, wings and breast). We therefore did not further develop a model for the total welfare score during the end-of-life stage, but will present the scores for both indicators separately. Similar to the analysis of the welfare score on-farm, the minimum, maximum, 10%, 50% and 90% percentiles were calculated and are presented in the results section, in addition to histograms presenting the distributions within and between production systems. Furthermore, Mood's median test and Wilcoxon signed rank tests were used to determine difference between the systems in scores for the indicators.

[1] This may change in the future and then lead to a different weighing per indicator

3 Results

3.1 Part 1: data collection for the four selected production systems

Tables 2-5 present the data on the level of the indicators, and also show the sample size and the source of the data per production stage for the four selected production systems. Likewise as for the tables on the justification of the choice of indicators in the report of de Jong [6], we start with the broiler table, followed by the broiler breeders, hatchery and end-of-life.

3.1.1 Broilers (on-farm)

Table 2 shows the data that could be collected for broiler welfare on-farm. Most data were routinely collected by the processing plant and were therefore collected for a large number of flocks over the year 2017 (Table 2). Only for the proportion of broilers resting on an elevated area no figures were available, but presence of elevated areas was also included in the enrichment scores. For locomotion, we used figures that have been collected for research purposes by one of the project partners, and these were from a small number of flocks. This implicates that these figures need to be interpreted with care as these have not been collected in a commercial setting, as for the other indicators.

Hardly any data were available for the organic production system, except a few data for footpad dermatitis (only based on 20 flocks) and total mortality, in addition to the stocking density and presence of enrichment/veranda or outdoor range/natural light and early feeding/water. Thus, overall, the welfare of the organic production system as compared to the other three production systems could not be evaluated. Organic flocks showed more mild footpad dermatitis as compared to the other systems, which was mainly due to higher scores in the winter season (data per season not shown).

In general, both the Dutch Retail Broiler and Better Life one star showed better values for the various welfare indicators than the conventional system, only the average proportion of chickens with scratches was higher for the Dutch Retail Broiler and Better Life one star than for conventional. Furthermore , averages for footpad dermatitis and hock burn were better for Better Life one star than for the Dutch Retail Broiler, although differences between these systems were smaller as compared to the differences between these two systems and conventional system. Mortality figures were more or less equal for all systems, but it should be taken into account that these are calculated over a longer rearing periods for the three systems with slower growing breeds (Dutch Retail Broiler, Better Life one star, organic) as compared to the conventional system. The proportion of flocks that received antibiotics treatment was much higher for conventional as compared to Dutch Retail Broiler and Better Life one star. Early feeding/water provision was more common for Dutch Retail Broiler and Better Life one star as compared to conventional. Stocking density was much lower for Better Life one star and organic as compared to the other two systems. These differences were present because of the different guidelines of the production systems. Only organic farms had an outdoor range, but no enrichment or natural light indoor. Only a relatively small proportion of conventional flocks had environmental enrichment if compared to Dutch Retail Broiler and Better Life one star, but more flocks had natural light in conventional than in Dutch Retail broiler. The latter is a consequence of comparing the systems for one supplier which does not include natural light in the Dutch Retail Broiler system. All flocks of Better Life one star had natural light and a veranda according to the guidelines of this system.

Table 2 *Welfare data of the broiler (on-farm) stage, according to the key-indicators as presented in Table 9 in [6] and collected in 2017 unless otherwise indicated. Data are presented as average (median-min-max) unless otherwise indicated in a footnote; N indicates the number of flocks sampled per production system. In case resource or management based indicators have been selected because of lack of data of the preferred animal-based indicator this is indicated in the column 'remarks'.*

Indicator	Conventional	Dutch Retail Broiler	Better Life one star	Organic	Remarks
Provision of early feeding and water					
% Flocks fed in the hatchery[1]	10	50	100	0	Alternative indicator for proportion of emaciated chickens
% On-farm hatched flocks[1]	7	0	No data	0	Alternative indicator for proportion of emaciated chickens
% broilers resting on an elevated structure	No data	No data	No data	No data	Elevated resting areas are also included in the enrichment score
Footpad dermatitis (0-2 scale)					
% score 1 (mild)	17.0 (14-0-94)	5.9 (2-0-87)	6.3 (2-0-66)	42.4 (41.5-8-75)	
% score 2 (severe)	18.6 (10-0-100)	3.5 (0-0-97)	2.3 (0-0-72)	3.5 (1-0-15)	
N	2743	2842	867	20 [2]	
Hock burn (0-1 scale)					
% score 1 (hock burn present)	12.9 (7.3-0-88.6)	8.2 (5.2-0-85)	3.8 (2.3-0-69)	No data	
N	2743	2843	867		
Breast irritation (0-1 scale)					
% score 1 (breast burn present)	0.3 (0-0-28.8)	0.2 (0-0-37.7)	0.2 (0-0-24)	No data	
N	2743	2843	867		
Maximum stocking density (at any moment in the production cycle), kg/m^2 [3]	39-42[4]	38	25	21	
Proportion of chickens with locomotion defects (0-5 scale) 5					
% score 1+2 (small defect)	77.5/87%	90/89.2%	86.7/89.1%	No data	
% score ≥3 (large defect)	20.8%/13%	4.2/5.8%	8.3/1.7%	No data	
N	4/7	4/4	4/4		
Total mortality, %	2.8 (2.5–0.1–19.2)	2.5 (2.1-0.1-19.8)	1.7 (1.4-0.1-12.3)	1.7[6]	
N	2505	2747	829		
Absence of curative antibiotics usage [7]					
% flocks without treatment	28.1	86.4	96.1	No data	
Scratches (0-1 score)					
% score 1 (scratches present)	1.0 (0.7-0-18)	2.1 (1.8-0-13.5)	1.8 (1.5-0-16.5)	No data	
N	2743	2843	867		
Presence of environmental enrichment[8]					Alternative indicator for % of time spent on species specific behaviours
% farms with 1 enrichment	0	100	0	0	
% farms with 2 enrichments	0	0	100	0	
% farms with 3 enrichments	19.3	0	0	0	
Presence of covered veranda/outdoor range					Alternative indicator for % of time spent on species specific behaviours

Indicator	Conventional	Dutch Retail Broiler	Better Life one star	Organic	Remarks
% farms with covered veranda	0	0	100	0	
% farms with outdoor range	0	0	0	100	
% Farms with natural daylight indoors 9	19.3	5	100	0	Alternative indicator for % of time spent on species specific behaviours

[1] Estimated proportions by the hatchery and processing plant; % on-farm hatched flocks based on number of chickens and not on number of flocks;
[2] sample of 20 flocks; data of 10 flocks collected in the winter season (March 2019) and data of 10 flocks collected in the summer season (August 2019), no other data available;
[3] stocking density according to legislation or producer guidelines or rules;
[4] maximum stocking density allowed differed between countries; part of the processed broilers are reared in Germany where a maximum stocking density of 39 kg/m^2 applies;
[5] no routinely collected data available; data from a trial in 2019 from a project partner in experimental (pens > 25 m^2) or semi-commercial conditions (annual average 2019 for 7 flocks of 12.000 broilers) for conventional; data for DRB, BLS vary dependent on the breed tested, therefore more than one figure is provided, and result from measures in experimental pens (pens > 25 m2) in 2019 by a project partner; all breeds were tested at a similar stocking density with 4 replicates per breed and 60 birds per pen scored for locomotion (replicate);
[6] estimation of the processing plant, no individual farm data provided;
[7] the processing plant registered the proportion of flocks with or without antibiotics treatments;
[8] figures provided by the processing plant and represent the proportion of farms (conventional) or according to rules or guidelines for the concept (Dutch Retail Broiler, Better Life one star, organic), enrichment comprises e.g. bales, pecking stones, perches, platforms, scattering grains;
[9] estimation of the processing plant (conventional, Dutch Retail Broiler) or concept guidelines/legal regulations (Better Life one star, organic); relates to natural light provided in the broiler house e.g. by roof windows or windows in the walls.

3.1.2 Broiler breeder stage (rearing and production)

Table 3 shows the welfare data of the broiler breeder stage, for both the rearing and production phase (if relevant). Footpad dermatitis and breast blister scores were provided by the processing plant, resulting in a large number of flocks of which data have been collected. For the other variables, data were provided by the hatchery and comprise a much lower number of flocks, or have been estimated by expert opinion of the project partners. For the 'Dutch Retail Broiler' two different breeds have been used, and if known, the values are presented or estimated for both breeds. Similar to the broiler on-farm stage, data for organic were lacking. Taken together, many of the data for the broiler breeder stage were either estimated, from a small number of flocks or not available, indicating that the estimation of overall welfare of parent stock has a high uncertainty.
Average footpad dermatitis scores and breast blister scores are higher for conventional than for parent stock of Dutch Retail Broiler and Better Life one star (for the latter only data from Hubbard flocks have been collected). Feed and water restriction are applied for conventional and to a lesser extent for Ranger flocks for Dutch Retail Broiler, whereas Hubbard dwarf females receive feed and water (nearly) ad libitum. Stocking density during rearing is slightly higher for flocks with Hubbard dwarf females in rearing, which are smaller if compared to conventional and Ranger females. Mortality during rearing and production is highest for conventional, with Ranger in between conventional and Hubbard flocks. Nearly all conventional breeders received antibiotics in rearing, in contrast to parent stock of slower growing broiler strains. Conventional production also had the highest number of flocks treated with antibiotics in production. In 2017, mutilations were performed in nearly all flocks. More farms with parent stock of slower growing broilers had environmental enrichment if compared to conventional parent stock. Unfortunately, data on resting on an elevated platform, injuries and locomotion were not available. Furthermore , several data were lacking for Ross Ranger flocks (Dutch Retail Broiler).

Table 3 *Welfare data of the broiler breeder stage, according to the key-indicators as presented in Table 10 in [6]. Data are presented as average (median-min-max) unless otherwise indicated in a footnote; N indicates the number of flocks sampled per production system. In case resource or management based indicators have been selected because of lack of data of the preferred animal-based indicator this is indicated in the column 'remarks'. For 'Dutch Retail Broiler', two different breeds have been used in practice and results may differ between breeds. Therefore, these data are presented in two separate columns.*

Indicator	Conventional	Dutch Retail Broiler (Ranger flocks)	Dutch Retail Broiler (Hubbard flocks (dwarf females))	Better Life one star	Organic	Remarks
Breed/line	Fast growing male/female	Slower growing female/fast growing male	Dwarf female/fast growing male	Dwarf female/fast growing male	Dwarf female/fast growing male	Alternative indicator for prevalence of stereotypic pecking
Water provision time						
Rearing, water control applied[1]	>3 weeks of age	>6 weeks of age	>8 weeks of age	>8 weeks of age	No data	
Production period, number of hours water provided [2]	6 h	12-14 h	12-14 h	12 – 14h	No data	
% breeders resting on an elevated structure (rearing+ laying)	No data	No data	No data	No data	No data	
Footpad dermatitis (1-5 scale)						
Average score	1.91 (1-1-5)	No data	1 (1-1-1)	1 (1-1-1)	No data	
N	680		28	14		
Breast blister (1-5 scale)		No data				
Average score	1.29 (1-1-5)		1.01 (1-1-1.2)	1 (1-1-1)	No data	
N	681		28	14		
Maximum stocking density (at any time) (rearing) birds/m^2 [3]	10 females/7 males	10 females/7 males	Dwarf females: 11 females/12 males	Dwarf females: 11 females/12 males	No data	
Maximum stocking density (at any time) (production) [4]	7 females and 0.6 males/m^2/6.9 birds/m^2	7 females and 0.6 males/m^2	7 dwarf females/m^2 and 0.6 males /6.9 birds/m^2	7 dwarf females/m^2 and 0.6 males	No data	
% breeders with locomotion defects	No data	No data	No data	No data	No data	
% total mortality (females) [5]						
Rearing	7	4	4	4	No data	
N						
Production	10/10.2	6	7/9.1	6	4.5	
N	180		8		4	
Absence of curative antimicrobials use						
Rearing, % flocks without antibiotic treatment (females only) [6]	53%	81%	81%	81%	81%	
Production, % flocks without antibiotic treatment [7]	81%	91%	91%	91%	91%	
N	180		8		4	
% breeders with injuries (scratches, wounds)	No data	No data	No data	No data	No data	
% flocks without mutilations [8]						
Females intact beak	100	1	1	1	No data	
Males intact beak	1	1	1	1	No data	

Indicator	Conventional	Dutch Retail Broiler	Dutch Retail Broiler	Better Life one star	Organic	Remarks
Males intact toe	1	1	1	1	No data	
Presence of environmental enrichment [9]						Alternative indicator for % of time spent on species specific behaviour
% farms with 1 enrichment	27.6	No data	0	0	No data	
% farms with 2 enrichments	41.4	No data	100	100	No data	
N (farms)	29		3			
Presence of covered veranda/outdoor range		0				Alternative indicator for % of time spent on species specific behaviour
% farms with covered veranda	0	0	0	0	No data	
% farms with outdoor range	0	0	0	0	No data	
N (farms)	29		3			
% farms with natural daylight	3.44/5	0	0/10	10	No data	Alternative indicator for % of time spent on species specific behaviour
N (farms)	29		3			

[1] Estimated age at which water restriction starts; estimation that this is applied in 90-95% of all flocks; estimation by the nutrition and breeding company;
[2] estimated time that water is available in the production period; 5-10% of the flocks receive unrestricted water supply; water available when lights on (to prevent leakage); estimation by the breeding and nutrition company;
[3] stocking density as indicated by the nutrition company;
[4] stocking density as indicated by the nutrition company, where two figures are provided, the first one is the estimation by the nutrition company and the second one is provided by the hatchery and the actual stocking density;
[5] for the rearing period, estimations as provided by the nutrition company, for the production period the first figure is the estimation provided by the nutrition company, the second figure is the average as provided by the hatchery; if only one figure is provided, it is an estimation by the nutrition company. The number of flocks relates to the hatchery figures;
[6] data as registered by Avined [12]; average annual daily dosage (add) 15.02, flocks with parent stock of slower growing strains add 4.53 (not separated per production system) [12];
[7] data for antibiotic treatments as registered by Avined [12]; conventional: (add) 3.58, parent stock of slower growing broiler strains, average add 0.68 [12] (not separated per production system);
[8] estimated by the nutrition and breeding company; males with intact beak/toe are test flocks;
[9] figures provided by the hatchery (where flock numbers are provided) or estimations from the nutrition company; for natural light if two figures are provided: first figure is a calculation from the hatchery and the second figure the estimation from the nutrition company.

3.1.3 Hatchery stage

Very few data were available for the hatchery stage. Only early feeding/water provision could be estimated by the hatchery and processing plant (Table 4). More flocks for Dutch Retail Broiler and Better Life one star received early feeding and water as compared to conventional and organic.

Table 4 *Welfare data of the broiler hatchery stage, according to the key-indicators as presented in Table 11 [6]. Data are presented as averages. In case resource or management based indicators have been selected (because of lack of data of the preferred animal-based indicator) this is indicated in the column 'remarks'.*

Indicator	Conventional	Dutch Retail Broiler	Better Life one star	Organic	Remarks
Proportion of flocks fed in the hatchery[1]	10	50	100	0	
Water provision in the hatchery[1]	10	50	100	0	
Proportion of flocks with on-farm hatching[1]	7	0	No data	0	Alternative indicator for proportion of chicks showing undisturbed resting; also for culling procedures, sexing, vaccinations and disinfection at the hatchery, and for proportion of chickens showing species specific behaviour
Cloacal temperature (hatchery and transport)	No data	No data	No data	No data	
Dead-on-arrival, %	No data	No data	No data	No data	

[1] Estimated proportion by the hatchery and the processing plant.

3.1.4 End-of-life stage

Table 5 shows the data for the end-of-life stage of broiler chickens (catching until processing, including transport). Very few data were available as only limited data are collected on a routine basis by the processing plant. Proportion of broilers dead-on-arrival decreases from conventional to organic. Breast and leg bruises are highest for Dutch Retail Broiler, whereas wing bruises are highest for conventional, followed by Better Life one star and then Dutch Retail broiler, for organic no figures are provided. All broilers have been stunned with gas and there was no conscious shackling for all production systems.

Table 5 *Welfare data of the broiler end-of-life stage, according to the key-indicators as presented in Table 12 in [6]. Data are presented as average (median-min-max) unless otherwise indicated in a footnote; N indicates the number of flocks sampled per housing system. In case resource or management based indicators have been selected (because of lack of data of the preferred animal-based indicator) this is indicated in the column 'remarks'.*

Indicator	Conventional	Dutch Retail Broiler	Better Life one star	Organic	Remarks
Body weight loss	No data	No data	No data	No data	Feed withdrawal time suggested as alternative indicator, but also no data available
Water withdrawal time (min)	No data	No data	No data	No data	
Stocking density in containers (kg/m^2)	No data	No data	No data	No data	Alternative indicator for behaviour in containers
% Dead-on-Arrival[1]	0.11 (0.08-0-1.26)	0.07 (0.04 – 0 - 4.27)	0.03 (0.02 – 0 – 1.62)	0 (0-0-0.32)	
N	2743	2840	866	341	
% broilers with trapped limbs	No data	No data	No data	No data	
% supine birds	No data	No data	No data	No data	
% broilers with bruises[1]					
Breast bruises	0.33 (0.2-0-8)	1.24 (1.2-0-21.3)	0.30 (0-0-4.7)	No data	
Leg bruises	0.29 (0.2-0-6)	1.01 (1-0-31)	0 (0-0-2)	No data	
Wing bruises	4.8 (4.6-0-78.0)	2.7 (2.7-0-25.0)	3.2 (3.0-0-12.5)	No data	
N	2743	2842	867	No data	
% broilers with splayed legs	No data	No data	No data	No data	
% broilers with wing fractures	No data	No data	No data	No data	
% broilers with leg fractures	No data	No data	No data	No data	
% flocks with conscious shackling[2]	0	0	0	0	
% flocks with gas stunning[2]	100	100	100	100	

[1] Data provided by the processing plant;
[2] relates to the processing plant involved in the project.

3.2 Part 2: welfare scores per indicator

3.2.1 Broiler welfare on-farm

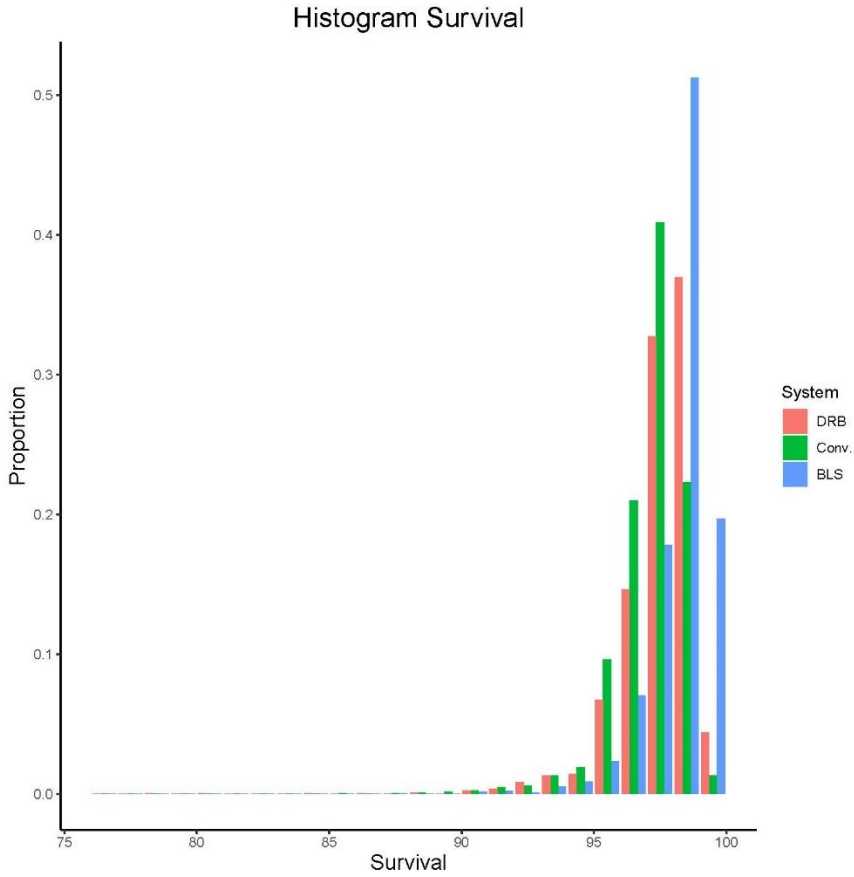
Table 6 presents the values and the calculated scores per welfare indicator for broilers on-farm on a scale from 0 to 100 per production system, and the variation in values and scores within the different production systems, over the period 2017-2018. Values and scores separated per year are presented in Appendix 7. Histograms presenting the distribution of the individual scores for the three production systems are shown in Figure 1. Scores for early feeding, stocking density and enrichment/natural light/veranda or outdoor range are not presented graphically, as these have only one value per production system or are expressed as percentage (early feeding).

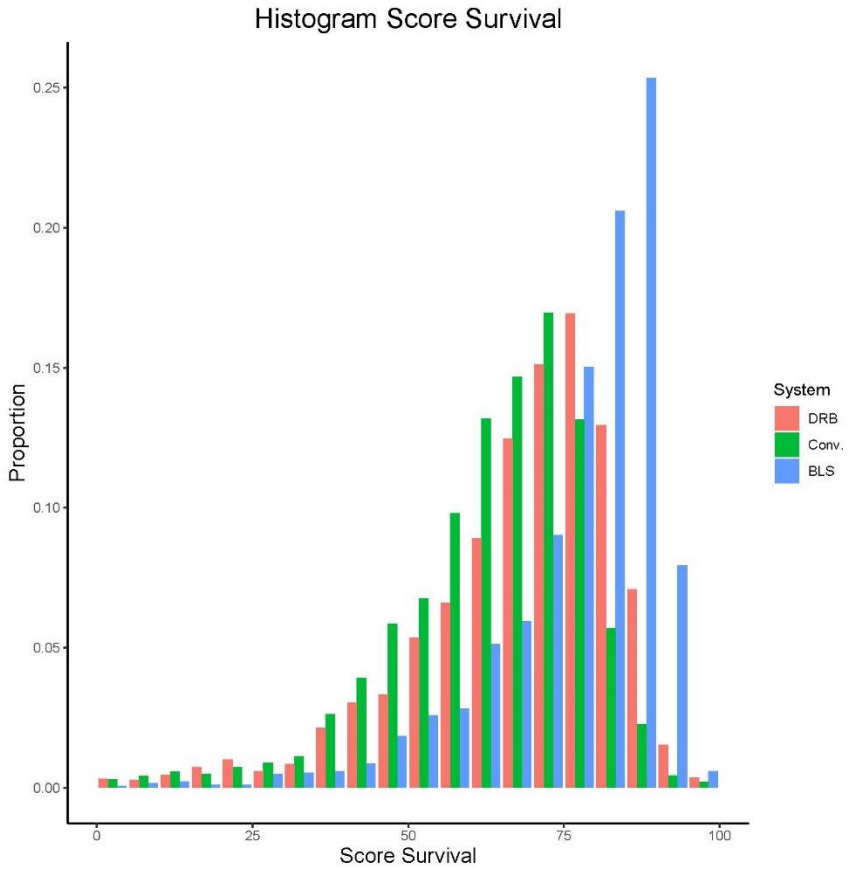
The tables in Appendix 6 show that differences between the production systems are similar for both years, i.e., Better Life one star having the best scores, followed by Dutch Retail Broiler and conventional. However, there are also some differences for some indicators within a production system across both years. This for example resulted in better total scores and total scores for animal based measures in 2018 than in 2017 for conventional and especially for Dutch Retail Broiler, whereas scores in 2017 and 2018 were more or less equal for Better Life one star. For mortality, Figure 1a and Table 6 indicate that Dutch Retail Broiler and Conventional have more or less similar scores, whereas Better Life one star has higher scores. A higher prevalence of footpad lesions and thus a lower score

was found for conventional versus the other systems. Clearly, conventional had a peak around 20 points whereas the other systems had a peak above 80 points, but for conventional the distribution over the range between 0-100 was more equal than for the other systems (Figure 1b and Table 6). Hock burn scores generally followed the same trend, with higher prevalence and lower score for Conventional than for the other systems, and the Better Life System having a peak around 100 points (no or very low prevalence of hock burn) (Figure 1c and Table 6). Prevalence of breast irritation was low in all systems resulting in more or less similar score distributions (Figure 1d and Table 6). Scratches were more prevalent in Better Life one star and Dutch Retail Broiler than in Conventional, resulting in a generally higher (better) score for Conventional than for the other production systems (Figure 1d and Table 6). All scores show that there is generally substantial overlap in scores between the systems, although the peak may be at a different level.

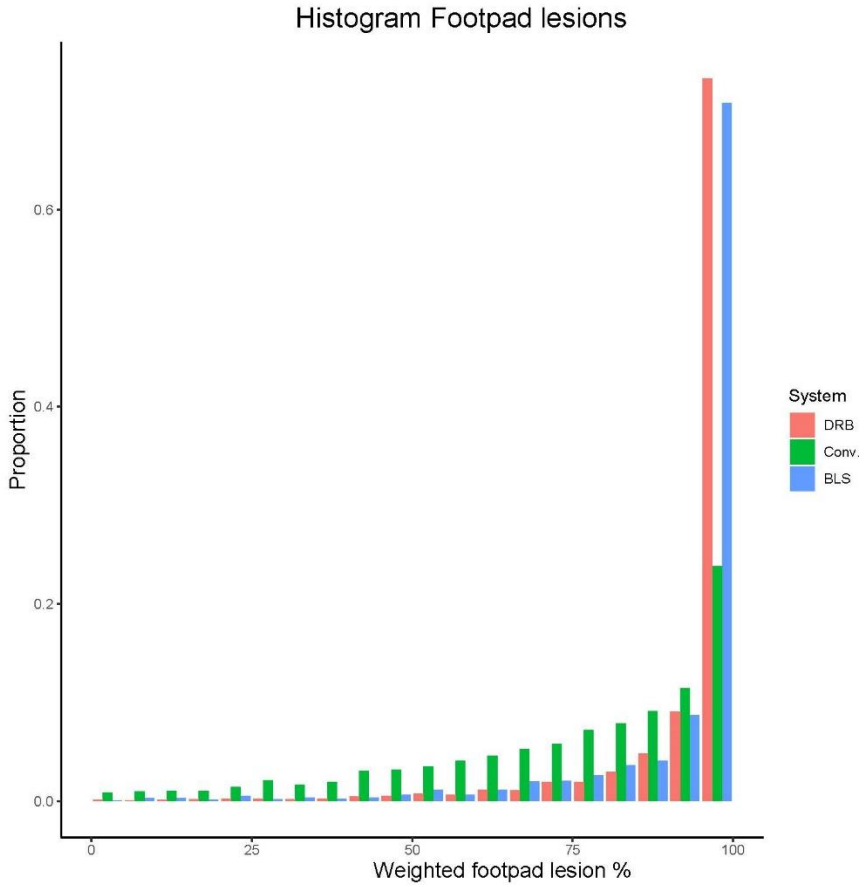
Table 7 summarises the results of the statistical analyses comparing the values and scores between the three production systems. Mood's test comparing the median scores and median values showed that for all indicators conventional significantly differed from Dutch Retail Broiler and Better Life one star. Comparing Dutch Retail Broiler and Better Life one star showed that the weighted fraction of footpad lesions and footpad lesion scores did not differ significantly, whereas the proportion of scratches differed but the score did not (i.e., experts did not assign different scores within this range of prevalence). Wilcoxon signed rank test generated more or less similar results, only the proportion of scratches did not differ between DRB and BLS when applying this test. Note that in comparing the systems most p-values are significant. Because the sample sizes are large, these tests have a large power to identify even small difference (and as a result, small differences are significant). The actual difference in median value (or some other measure) is thus likely more informative than the p-value. Appendix 7 presents the p-values of the statistical comparison between systems for 2017 and 2018 separately.

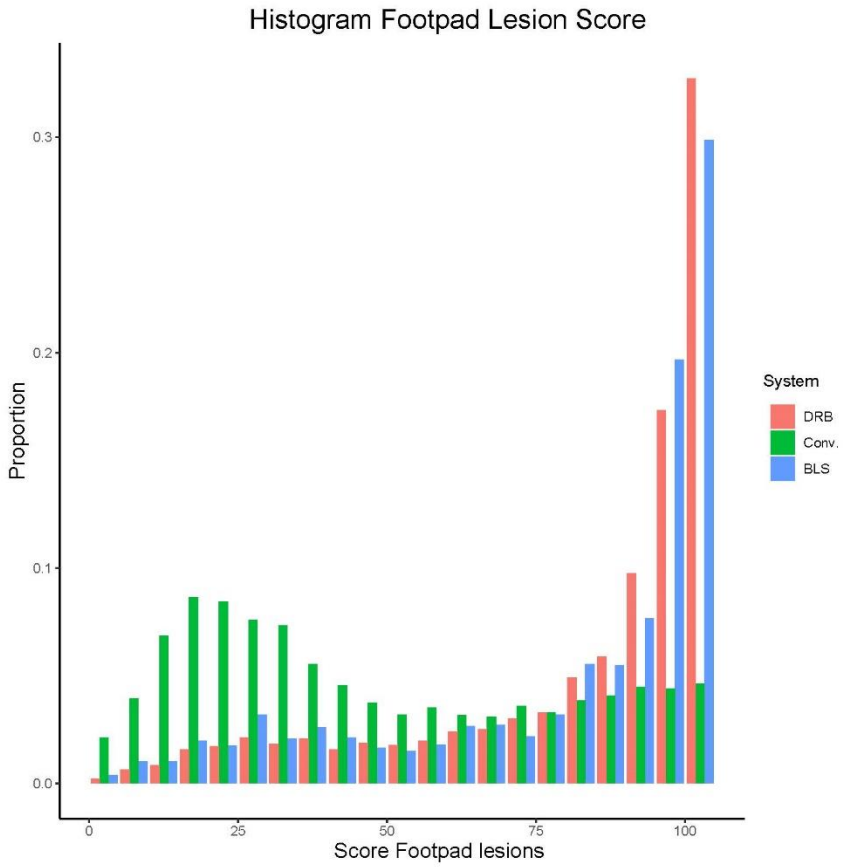
Locomotion (gait score) is not measured on a routine basis and gait scores are thus not included in Table 6 and 7 and in the figures. Appendix 6 presents data collected on locomotion by one of the project partners, and the associated scores, as an example. These data are however too limited to be included in the overall welfare score per production system.



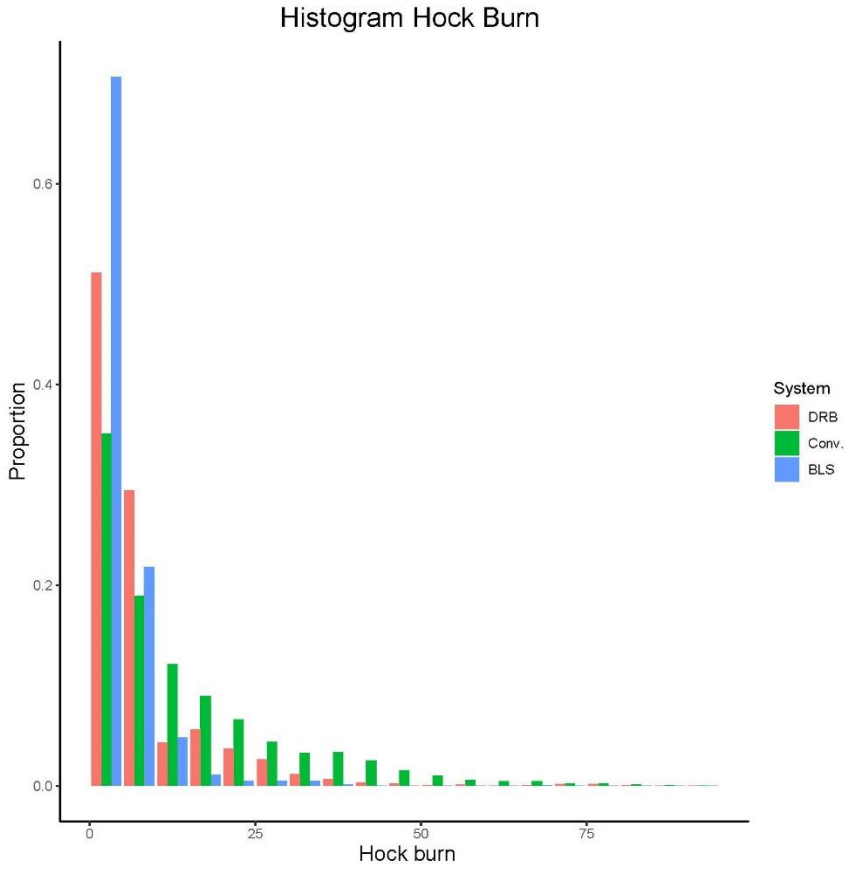


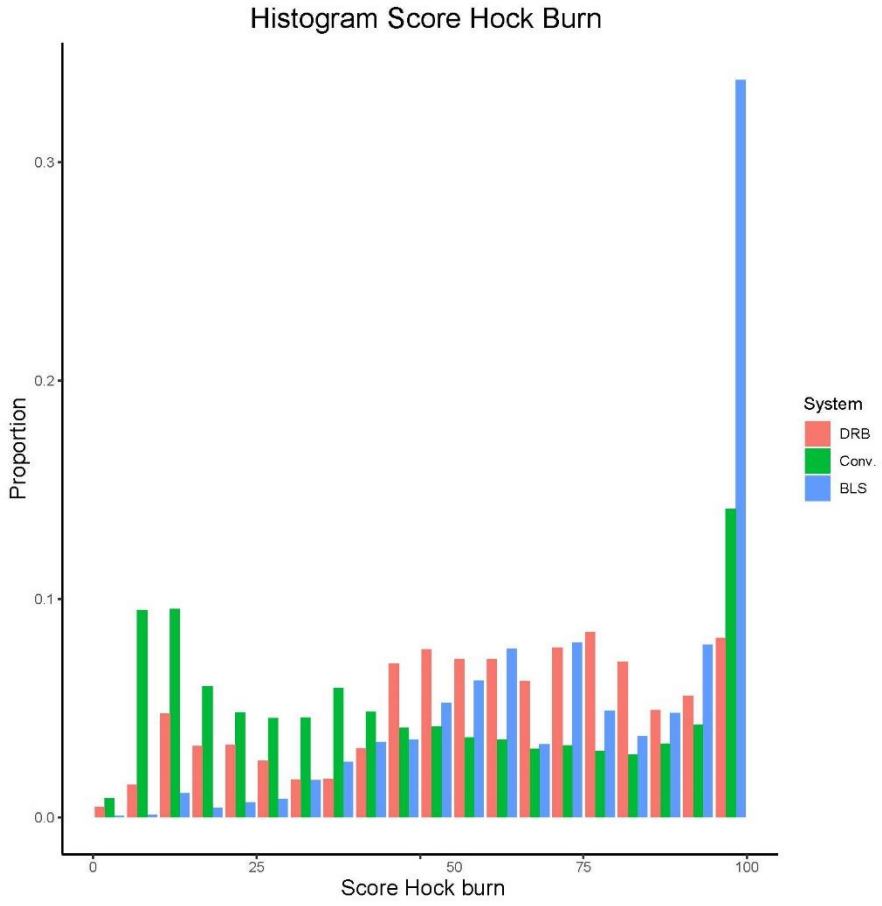
a. Survival (100%-mortality%)



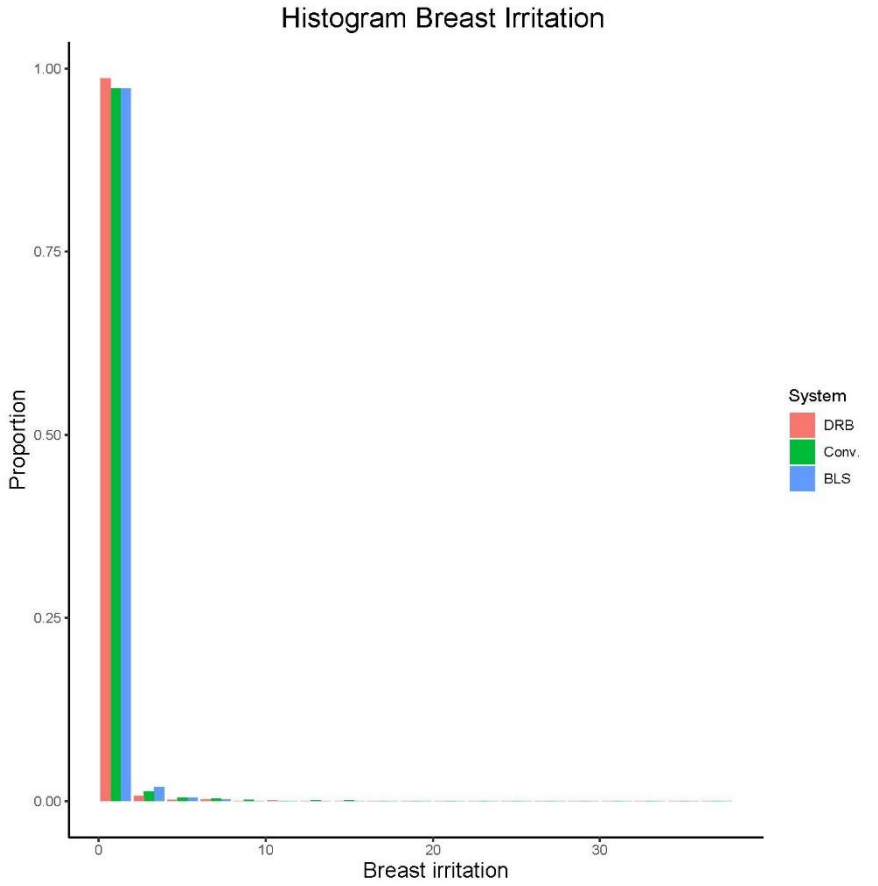


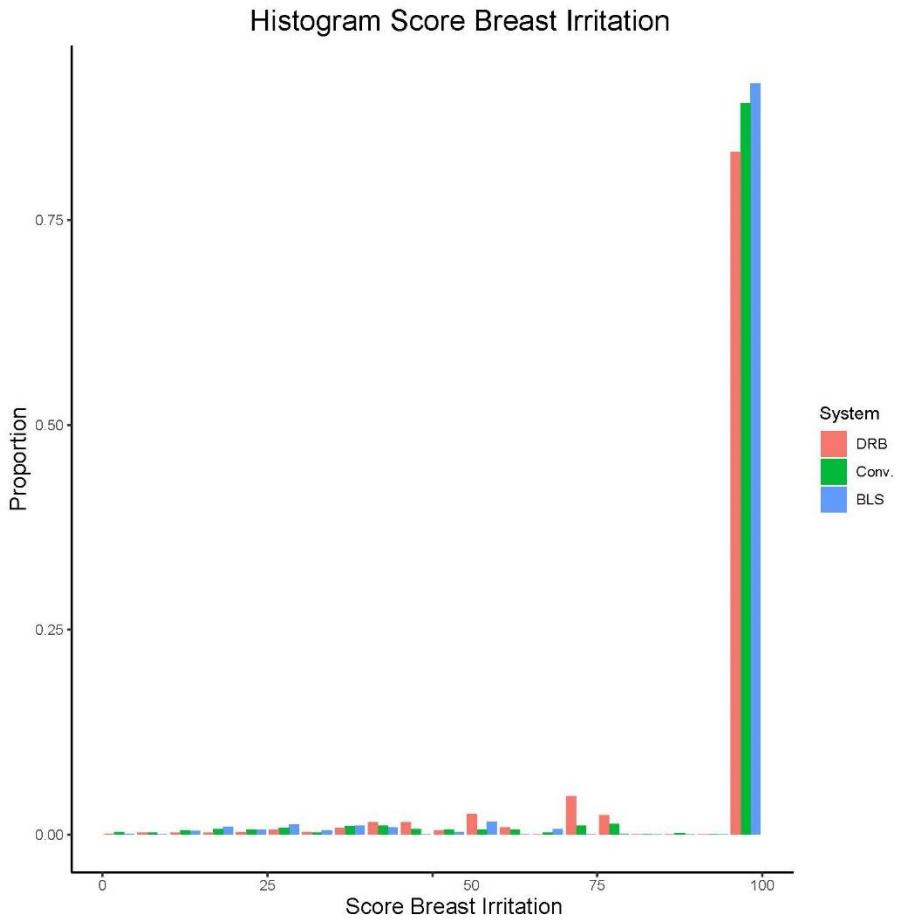
b. Footpad lesions. Weighted footpad lesion % = 100−(2*score1+7*score2)/7


Histogram Hock Burn
Proportion
0.0
0.2
0.4
0.6
0
25
50
75
Hock burn
System
DRB
Conv.
BLS

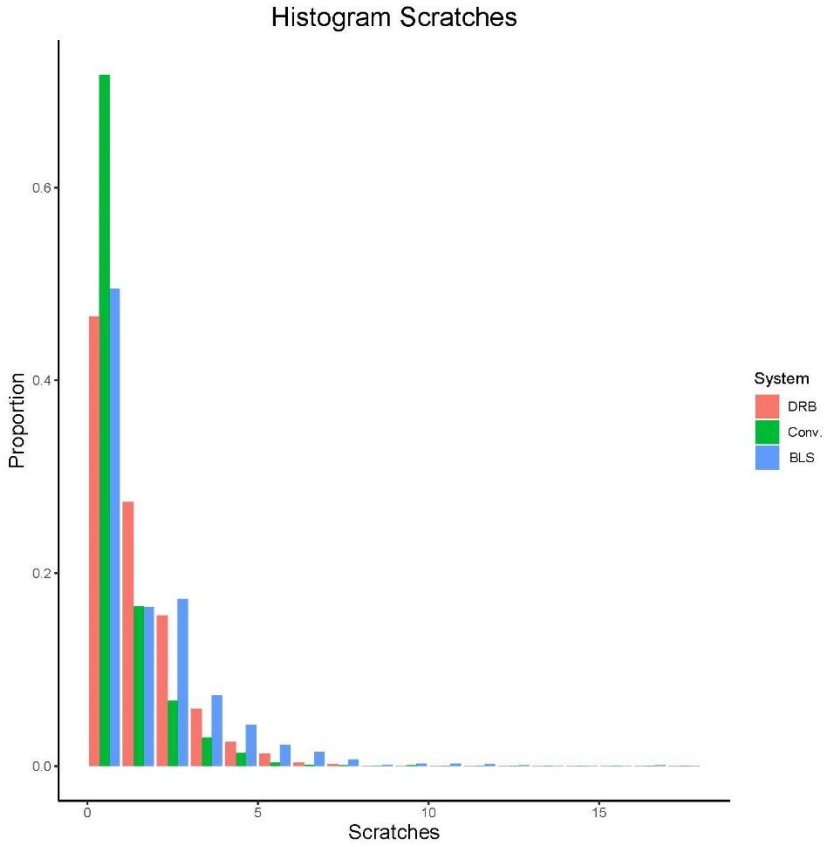

Histogram Score Hock Burn
Proportion
0.0
0.1
0.2
0.3
0
25
50
75
100
Score Hock burn
System
DRB
Conv.
BLS

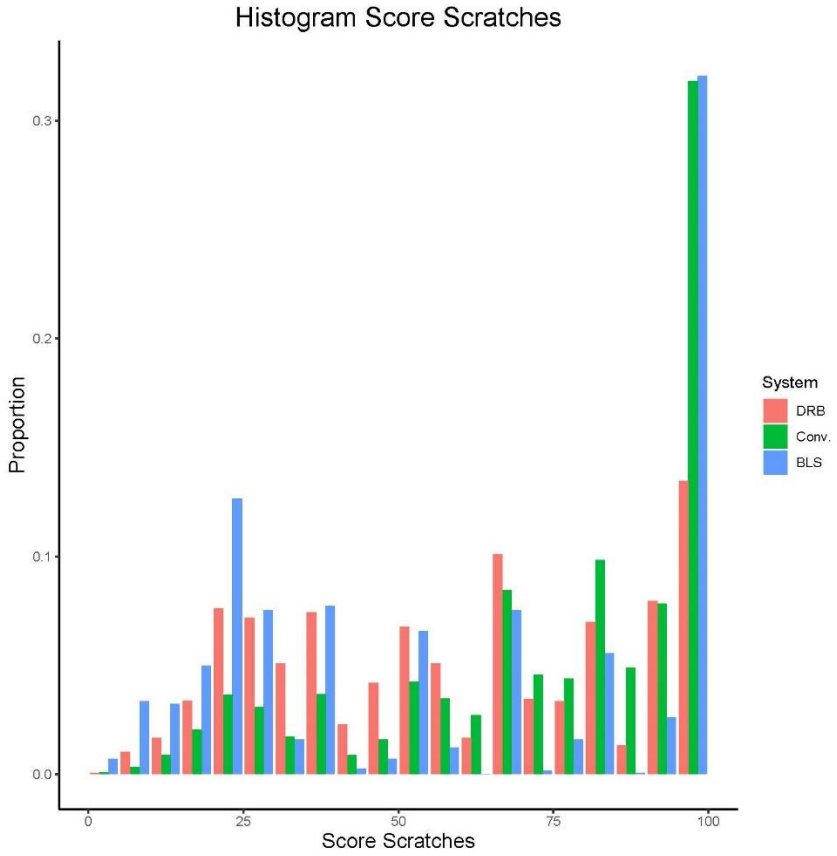
c. Hock burn.





d. Breast irritation





e. Scratches

Figure 1 *Histograms presenting the distribution of prevalence and scores for survival (100-mortality) (a), footpad dermatitis (b), hock burn (c), breast irritation (d), and scratches (e) for flocks of the three production systems: DRB (Dutch Retail broiler), Conv. (conventional) and BLS (Better Life one star).*

Table 6 *Values and calculated welfare scores for the broiler (on-farm) stage, based on data collected in 2017 and 2018. Values and calculated scores are presented as the minimum and maximum value or score, and the 0.10, 0.50 and 0.90 percentile to illustrate the distribution of the scores. Table A presents the results for the conventional farms, Table B for the Dutch Retail Broiler and Table C for the Better Life one star production system. Note, that for the values of footpad dermatitis, hock burn, breast irritation and scratches these are presented from 'best' to 'worst', whereas for the scores, this is the other way round, i.e. from 'worst' to 'best' score.*

A. Conventional (N=5683 flocks)

Variable and score	minimum	P0.10	P0.50	P0.9	maximum
Survival %	76.65	95.60	97.40	98.38	99.85
Mortality score	2.21	46.50	66.18	79.30	98.19
Weighted % footpad lesions[1]	0	37.43	81.43	98.86	100
Footpad lesion score	0	13.09	39.45	93.93	100
Hock burn %	0	0.5	9	36	88.6
Hock burn score	2.76	9.89	44.35	98.10	100
Breast irritation %	0	0	0	0.2	30
Breast burn score	0.44	78.37	100	100	100
Scratches %	0	0	0.5	2.5	18
Scratches score	1.42	29.17	84.86	100	100
Stocking density, kg/m^2 [2]			42		
Stocking density score[2]			0		
Hatchery fed, % flocks			0		
On-farm hatched, % flocks			0.7		
Early feeding score	25	25	25	25	81.67
1 Enrichment, % flocks[2]			0		
2 Enrichments, % flocks[2]			0		
3 Enrichments, % flocks[2]			0		
Daylight entrance[2, 3]			0		
Veranda[2, 3]			0		
Outdoor range[2, 3]			0		
Enrichment score[2]			18.25		

[1] First, fractions with score 0, 1 and 2 were calculated. Then, the weighted score was calculated according to the formula: 100 - (2 * fraction score1 * 100 + 7 * fraction score2 * 100)/7 (see [7] page 45). [2] For these variables, only one value was assigned to the system, which is presented in the column 'median'. [3] 0=not present, 1=present.

B. Dutch Retail Broiler (N=5936 flocks)

Variable and score	minimum	P0.10	P0.50	P0.9	Maximum
Survival %	76.15	95.84	97.79	98.77	99.99
Mortality score	2.20	46.30	71.40	84.47	99.87
Weighted % footpad lesions [1]	0	80.34	99.14	100	100
Footpad lesion score	0	37.86	95.43	100	100
Hock burn %	0	1	5	19.5	92.5
Hock burn score	2.43	20.14	64.60	94.56	100
Breast irritation %	0	0	0	0.3	37.7
Breast burn score	0.02	70.37	100	100	100

Variable and score	minimum	P0.10	P0.50	P0.9	Maximum
Scratches %	0	0	1.3	3.2	13.5
Scratches score	2.72	23.55	57.41	100	100
Stocking density, kg/m^2 [2]			38		
Stocking density score[2]			22.32		
Hatchery fed, % flocks			1.45		
On-farm hatched, % flocks			0.42		
Early feeding score	25	25	25	25	81.67
1 Enrichment, % flocks[2]			100		
2 Enrichments, % flocks[2]			0		
3 Enrichments, % flocks[2]			0		
Daylight entrance[2, 3]			0		
Veranda[2, 3]			0		
Outdoor range[2, 3]			0		
Enrichment score			29.1		

[1] First, fractions with score 0, 1 and 2 were calculated. Then, the weighted score was calculated according to the formula: 100 - (2 * fraction score1 * 100 + 7 * fraction score2 * 100)/7 (see [7] page 45). [2] For these variables, only one value was assigned to the system, which is presented in the column 'maximum'. [3] 0=not present, 1=present.

C. Better Life one star (n=1889 flocks)

Variable and score	minimum	P0.10	P0.50	P0.9	Maximum
Survival %	87.69	96.9	98.51	99.17	99.93
Mortality score	2.97	59.58	81.10	89.68	99.16
Weighted % footpad lesions[1]	4.71	74.28	98.86	100	100
Footpad lesion score	2.24	31.19	93.93	100	100
Hock burn %	0	0	2.67	9	69
Hock burn score	4.82	44.35	80.87	100	100
Breast irritation %	0	0	0	0	24
Breast burn score	1.13	100	100	100	100
Scratches %	0	0	1.33	4	16.5
Scratches score	1.71	18.72	56.38	100	100
Stocking density, kg/m^2 [2]			25		
Stocking density score[2]			57.36		
Hatchery fed, % flocks			100		
On-farm hatched, % flocks			0		
Early feeding score			76.67		
1 Enrichment, % flocks[2]			0		0
2 Enrichments, % flocks[2]			100		100
3 Enrichments, % flocks[2]			0		0
Daylight entrance[2, 3]			1		Yes
Veranda[2, 3]			1		Yes
Outdoor range[2, 3]			0		0
Enrichment score[2]			72.25		

[1] First, fractions with score 0, 1 and 2 were calculated. Then, the weighted score was calculated according to the formula: 100 - (2 * fraction score1 * 100 + 7 * fraction score2 * 100)/7 (see [7] page 45). [2] For these variables, only one value was assigned to the system, which is presented in the column 'maximum'. [3] 0=not present, 1=present.

Table 7 *Summary of results of statistical analysis comparing the indicator scores between the different systems. Results of the Mood's analysis are based on comparison of the median values of the*

scores, the Wilcoxon rank-sum test compares the sum of ranks of the systems and tests whether the systems have a similar distribution. The table shows the P-values of the comparison between two systems. Data from 2017 and 2018 have been included. Note: enrichment and stocking density scores have not been tested, as these scores were similar between flocks within a particular system.

	Conventional vs Dutch Retail Broiler		Conventional vs Better Life one star		Dutch Retail Broiler versus Better Life one star	
Indicator	**P-value Mood**	**P-value Wilcoxon**	**P-value Mood**	**P-value Wilcoxon**	**P-value Mood**	**P-value Wilcoxon**
Survival%	<0.0001	<0.0001	<0.0001	0	<0.0001	<0.0001
Mortality score	<0.0001	<0.0001	<0.0001	0	<0.0001	<0.0001
Weighted % footpad lesions[1]	0	0	<0.0001	0	0.7191	0.0251
Footpad lesion score	0	0	<0.0001	0	0.7191	0.0251
Hock burn %	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Hock burn score	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Breast burn %	<0.0001	<0.0001	0.0027	0.0049	<0.0001	<0.0001
Breast burn score	0.0189	<0.0001	0	0.0049	0	<0.0001
Scratches %	<0.0001	<0.0001	<0.0001	<0.0001	0.0072	0.1191
Scratches score	<0.0001	<0.0001	<0.0001	<0.0001	0.3290	0.1191
Early feeding score	<0.0001	<0.0001	0	0	0	0

[1] First, fractions with score 0, 1 and 2 were calculated. Then, the weighted score was calculated according to the formula:
100 - (2 * fraction score1 * 100 + 7 * fraction score2 * 100)/7 (see [7] page 45).

3.2.2 Broiler welfare during the end-of-life stage

Table 8 presents the values and the calculated welfare scores for Dead-on-Arrival (DoA) and the proportion of bruises per production system, and the variation in values and scores within the different production systems over the period 2017-2018. Values and scores separated per year are presented in Appendix 7. Histograms presenting the distribution of the individual scores for the three production systems are shown in Figure 2. The tables in Appendix 6 show that differences between both years were very small. Figure 2a and Table 8 show that Dead-on-Arrival was higher, and thus scores were lower, in conventional than in Better Life one star, with Dutch Retail Broiler in-between. Figure 2b and Table 8 shows that bruises were more prevalent in Conventional and Dutch Retail Broiler than in Better Life one star, resulting in higher (better) scores for Better Life one star than for the other two systems. Table 9 shows the results of the comparison of values scores for DoA and bruises between the production systems. Both values and scores differ significantly in both statistical comparisons. Appendix 7 also presents the statistical comparison of values and scores per year.

Table 8 *Values and calculated welfare scores for the end-of-life stage, based on data collected in 2017 and 2018. Values and calculated scores are presented as the minimum and maximum value or score, and the 0.10, 0.50 and 0.90 percentile to illustrate the distribution of the scores. Table A presents the results for the conventional farms, Table B for the Dutch Retail Broiler and Table C for the Better Life one star production system. Note, that for the values these are presented from 'best' to 'worst', whereas for the scores, this is the other way round, i.e. from 'worst' to 'best' score.*

A. Conventional (N=5683 flocks)

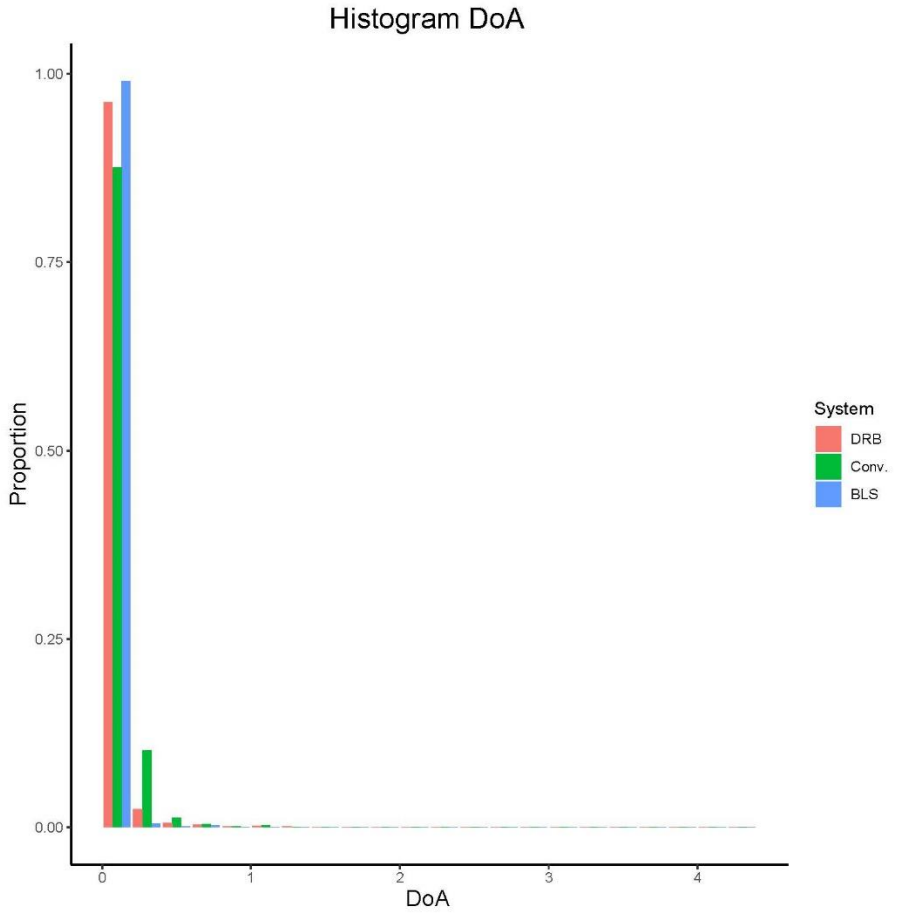
Indicator and score	minimum	P0.10	P0.50	P0.9	Maximum
DoA	0	0.02	0.08	0.22	3.87
DoA score	0.44	30.01	51.40	82.11	100
Bruises (total %)	0	2	5	8.1	29.5
Bruises score	0	16.10	42.22	74	100

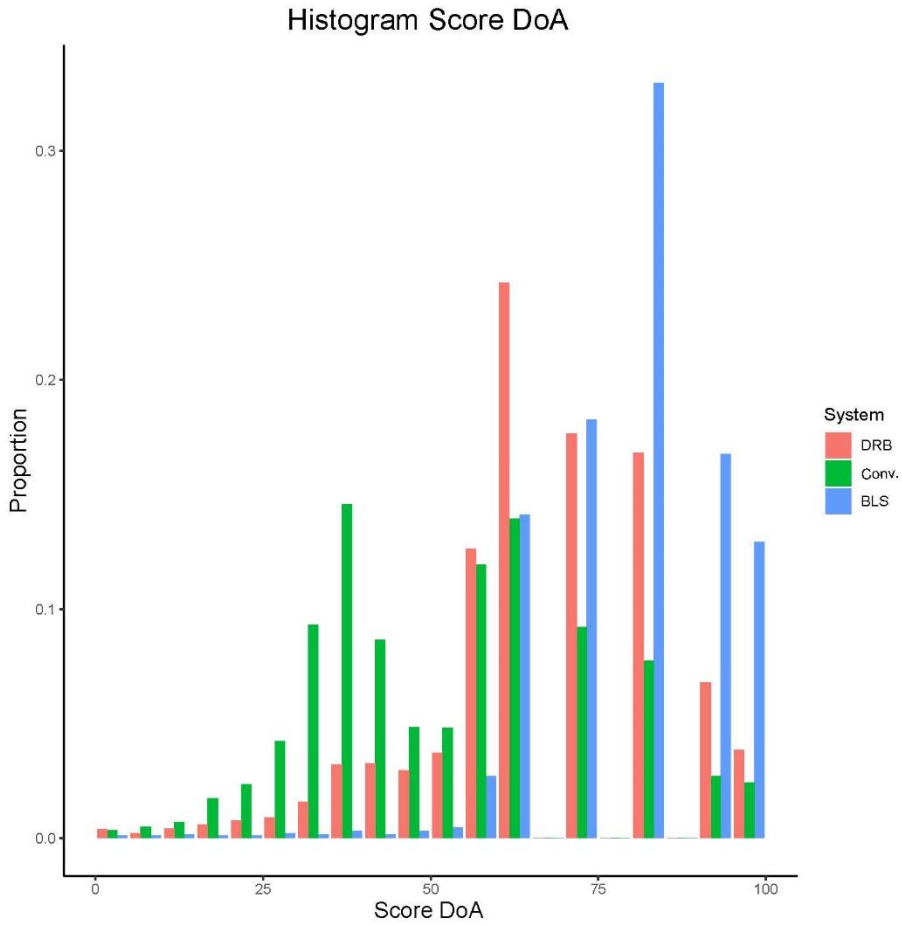
B. Dutch Retail Broiler (N=5936 flocks)

Indicator and score	minimum	P0.10	P0.50	P0.9	Maximum
DoA	0	0.01	0.04	0.10	4.27
DoA score	0.23	44.71	64.78	90.22	100
Bruises (total %)	0	3	4.8	6.7	26
Bruises score	0.24	27.72	43.81	61	100

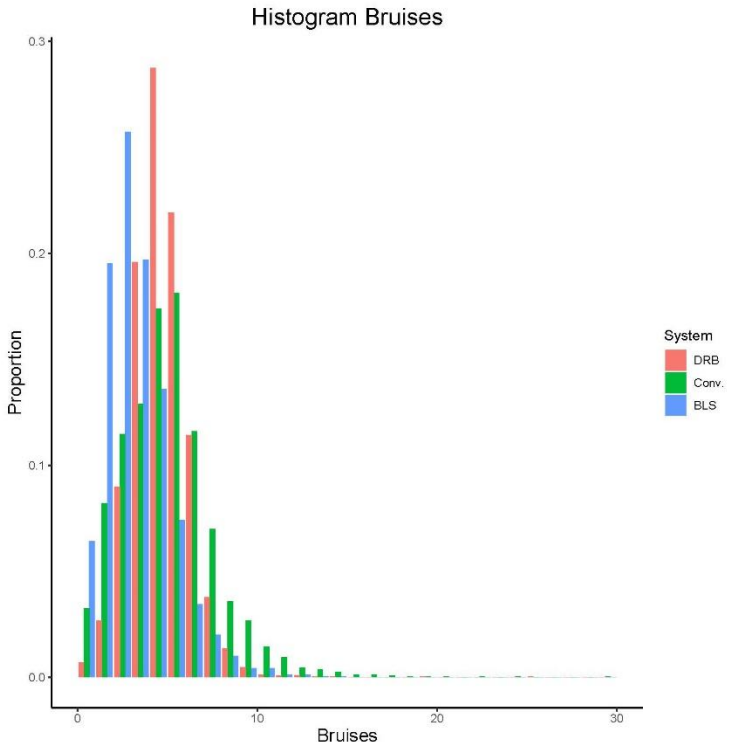
C. Better Life one star (N=1889 flocks)

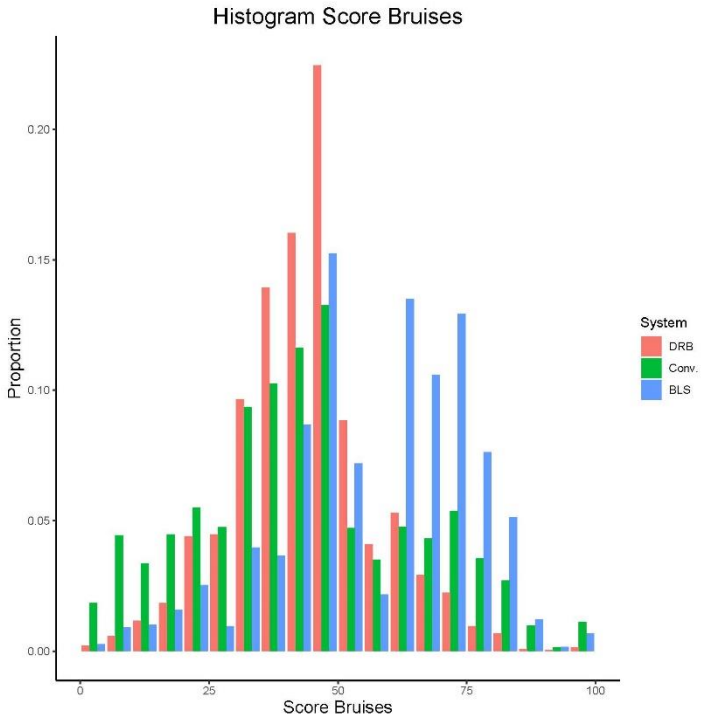
Indicator and score	minimum	P0.10	P0.50	P0.9	Maximum
DoA	0	0	0.02	0.04	2.57
DoA score	1.54	64.78	82.11	100	100
Bruises (total %)	0	1.5	3	6	14.5
Bruises score	3.55	33.91	61	79.12	100





a. Dead-on-Arrival





b. Bruises

Figure 2 *Histograms of the distribution of the prevalence and scores for Dead-on-Arrival (a) and Bruises (b) for flocks of the three production systems: DRB (Dutch Retail broiler), Conv. (conventional) and BLS (Better Life one star).*

Table 9 *Summary of results of statistical analysis comparing the indicator scores between the different systems for Dead-on-Arrival and bruises. Results of the Mood's analysis are based on comparison of the median values of the scores, the Wilcoxon rank-sum test compares the sum of ranks of the systems and tests whether the systems have a similar distribution. The table shows the P-values of the comparison between two systems. Data from 2017 and 2018 have been included.*

	Conventional vs Dutch Retail Broiler		Conventional vs Better Life one star		Dutch Retail Broiler versus Better Life one star	
Indicator	P-value Mood	P-value Wilcoxon	P-value Mood	P-value Wilcoxon	P-value Mood	P-value Wilcoxon
Dead-on-Arrival, %	<0.0001	0	0	0	<0.0001	<0.0001
Dead-on-Arrival, score	<0.0001	0	0	0	<0.0001	<0.0001
Total bruises, %	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Total bruises, score	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001

3.2.3 Overall welfare score calculation

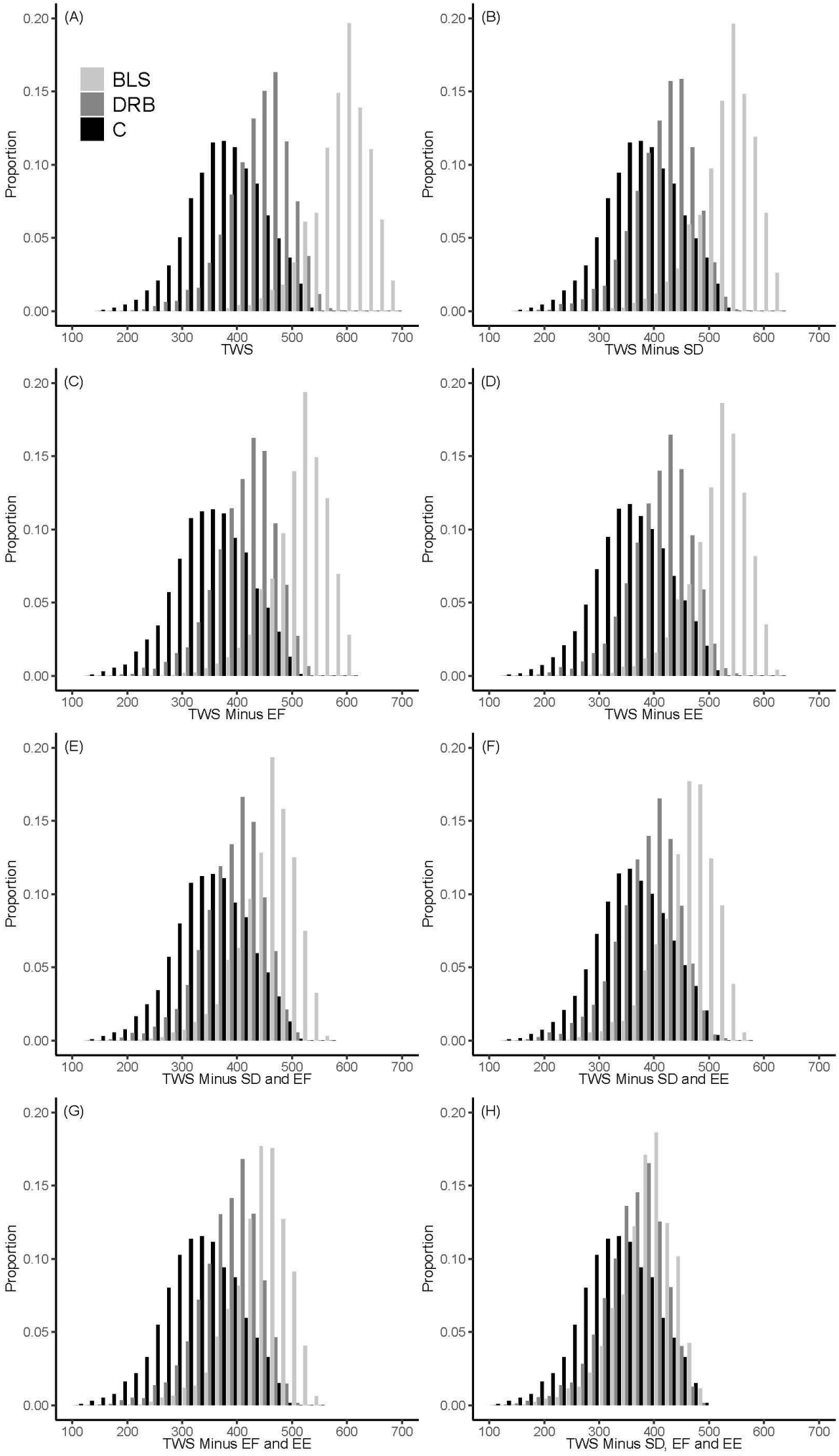
3.2.3.1 Correlation coefficients between indicators

Table 10 presents the correlation coefficients between all indicators as recorded by the processing plant, and separated per production system. This analysis shows that there is only one moderate correlation for all three systems, i.e., between footpad dermatitis and hock burn, and only for Conventional and Dutch Retail Broiler, and that all other correlations are very low. Based on these correlations, we decided that a simplification of the final welfare model, i.e. omitting one indicator or predicting one variable from another indicator, would not be possible. This thus means that all indicators were included in the calculation of overall welfare score.

3.2.3.2 Overall welfare score for broilers on-farm

To generate an overall on-farm welfare score, individual scores of the indicators were summed. The maximum score that could be achieved was 800 points, as 8 indicators were included. This resulted in substantial variation in scores within and between flocks in the three production systems (Figure 3A).

Figure 3 (page 36) *Histogram of the distribution of the Total Welfare Score (TWS) for flocks of the three production systems (histogram A). The TWS for a flock is the sum of the eight scores for the individual welfare measures. The maximum and best score that can be received is 800. Histograms B-H present the results of the sensitivity analysis and show the TWS minus the scores for Stocking Density (SD) (Histogram B), Early Feeding (EF) (Histogram C) and Environmental Enrichment (EE) (Histogram E) separately, and combined (F, G, H). Thus, Histogram H is based on only the 5 animal-based welfare indicators and each flock can therefore only be assigned a score between 0-500.*


(A)
BLS
DRB
C
Proportion
TWS
(B)
TWS Minus SD
(C)
TWS Minus EF
(D)
TWS Minus EE
(E)
TWS Minus SD and EF
(F)
TWS Minus SD and EE
(G)
TWS Minus EF and EE
(H)
TWS Minus SD, EF and EE

However, in the current model three resource-based indicators were included, and the respective scores for these indicators were similar for flocks within production systems for stocking density and enrichment, and variation was low for scores for early feeding. In the ideal situation, these will be replaced by animal-based measures that will likely result in much higher variation between flocks within a certain production system. Therefore, a sensitivity analysis was performed by omitting these resource-based measures one-by-one or in pairs; results are shown in Figure 3B-H.

Figure 3 illustrates that the three resource-based indicators that have been included in the current model have a relatively large effect on the total welfare score. Omitting these on-by-one results in a distribution of scores for the production system that resembles the total welfare score in Figure 3, omitting these in pairs results in a higher overlap in scores between the production systems, and omitting all three resource-based indicators results in a considerable higher overlap in scores between the three production systems. Table 11 shows the percentiles, minimum and maximum scores per production system for both the Total Welfare Score (8 variables) as the score based on animal based measures only (Total Score_ABM, 5 variables). It also illustrates the relatively high impact on the total score of these resource-based measures. Interestingly, for the Total score_ABM, comparing conventional with DRB shows that the minimum and maximum values are more or less equal, but that for DRB more flocks have a high score (0.5 percentile is 40 points higher than for conventional), whereas comparing DRB with BLS shows that the minimum level of BLS is higher than DRB and more flocks receive a high score, indicated by the higher median and 0.9 percentile. The statistical comparison of scores still showed a significant difference between the production systems, both for the Total Welfare Score as for the Total Score_ABM (Table 12).

Table 10 *Correlation coefficients between all variables recorded by the processing plant for the conventional (A), Dutch Retail Broiler (B) and Better Life one start system (C) separately. Moderate correlations (r^2>0.3) are indicated in bold.*

A. Conventional

Variable	Total % bruises	Hock burn	Breast irritation	Scratches	Footpad dermatitis[1]	Mortality
Dead-on-Arrival	0.026	-0.11	-0.0037	-0.23	0.029	-0.0046
Total % bruises		0.13	0.094	0.20	0.04	0.049
Hock burn			0.24	0.023	**0.44**	0.086
Breast Irritation				0.051	0.18	0.034
Scratches					-0.05	0.044
Footpad dermatitis						-0.072

[1] First, fractions with score 0, 1 and 2 were calculated. Then, the weighted score was calculated according to the formula: 100 - (2 * fraction score1 * 100 + 7 * fraction score2 * 100)/7 (see [7] page 45).

B. Dutch Retail Broiler

Variable	Total % bruises	Hock burn	Breast irritation	Scratches	Footpad dermatitis[1]	Mortality
Dead-on-Arrival	0.055	-0.04	0.062	0.13	-0.072	0.086
Total % bruises		0.13	0.062	0.012	0.076	-0.015
Hock burn			0.20	-0.15	**0.44**	-0.095
Breast Irritation				0.13	0.12	-0.014
Scratches					-0.16	-0.012
Footpad dermatitis						-0.088

[1] First, fractions with score 0, 1 and 2 were calculated. Then, the weighted score was calculated according to the formula: 100 - (2 * fraction score1 * 100 + 7 * fraction score2 * 100)/7 (see [7] page 45).

C. Better Life one star

Variable	Total % bruises	Hock burn	Breast irritation	Scratches	Footpad dermatitis[1]	Mortality
Dead-on-Arrival	0.13	-0.10	0.036	0.26	-0.044	0.036
Total % bruises		0.089	0.13	0.29	0.014	-0.0088
Hock burn			0.11	-0.34	0.27	-0.076
Breast Irritation				0.098	0.063	-0.028
Scratches					-0.2	0.057
Footpad dermatitis						-0.033

[1] First, fractions with score 0, 1 and 2 were calculated. Then, the weighted score was calculated according to the formula: 100 - (2 * fraction score1 * 100 + 7 * fraction score2 * 100)/7 (see [7] page 45).

Table 11 *Calculated total welfare scores for the broiler (on-farm) stage, based on data collected in 2017 and 2018. Scores are presented as the minimum and maximum score, and the 0.10, 0.50 and 0.90 percentile to illustrate the distribution of the scores, for the three production systems. Total welfare score is the score including all eight indicators, Total Score_ABM is the sum of scores excluding stocking density, early feeding and environmental enrichment. Note that the maximum score that can be achieved for a flock is 800 for the Total Welfare Score and 500 for the Total Score_ABM.*

Production system	N flocks	Variable	Minimum	P0.10	P0.50	P0.9	maximum
Conventional	5683						
		Total welfare score	155.40	288.32	374.40	462.08	531.60
		Total score_ABM	112.15	245.01	330.84	418.34	488.35
Dutch Retail Broiler	5936						
		Total welfare score	201.60	381.01	460.09	517.80	598.46
		Total score_ABM	113.04	291.93	370.87	427.25	486.94
Better Life one star	1889						
		Total welfare score	357.60	526.50	602.38	655.83	696.25
		Total score_ABM	151.32	320.22	396.10	449.55	489.97

Table 12 *Summary of results of statistical analysis comparing the Total Welfare Score and Total Score_ABM between the different systems. Results of the Mood's analysis are based on comparison of the median values of the scores, the Wilcoxon rank-sum test compares the sum of ranks of the systems and tests whether the systems have a similar distribution. The table shows the P-values of the comparison between two systems. Data from 2017 and 2018 have been included.*

	Conventional vs Dutch Retail Broiler		Conventional vs Better Life one star		Dutch Retail Broiler versus Better Life one star	
Indicator	P-value Mood	P-value Wilcoxon	P-value Mood	P-value Wilcoxon	P-value Mood	P-value Wilcoxon
Total Welfare Score	0	0	0	0	0	0
Total Score_ABM	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001

4 Discussion and conclusions

4.1 Part 1: data of all production stages for the different production systems

Firstly, we tried to collect as many data as possible for the welfare indicators of the four different phases of the broiler production chain. The tables in the results section show that most data are routinely collected for the broiler on-farm phase, followed by the broiler breeder phase, and that relatively little data are routinely collected for the end-of-life phase and the hatchery phase respectively. Furthermore, a lot of effort had to be paid in collecting the data for the breeder phase. The broiler breeder table includes several estimations, showing that there is a need for standardised collection of data by the different chain partners. Data are also collected by different chain partners (processing plant, nutrition company and hatchery) which makes data collection more complicated and maybe less comparable when only one partner collects the data. With respect to scoring welfare indicators at the plant, the scoring system of the processing plant for broiler breeders differs from that for broiler chickens in the sense that one score is assigned to a flock for e.g. footpad dermatitis, whereas for broiler chickens a 100 feet are scored in a standardised way. This makes it difficult to compare scores and it also involves a higher degree of subjectivity.

Only very few data could be collected for organic flocks. This is due to the fact that the organic flocks of our partner in Greenwell are processed abroad, and data are not collected on a routine basis. Therefore, the welfare of organic flocks cannot be estimated compared to the other production systems (conventional, Dutch Retail Broiler and Better Life one star). Earlier attempts to collect data from organic flocks also resulted in very few data on the actual welfare status [15, 16]. As far as we know, there are also only few data available in scientific literature on welfare performance of organic broiler flocks (France: [17]; Sweden: [18], but under experimental conditions and without outdoor range) and because housing conditions also differ between countries, extrapolation of these data to other countries is not possible. It would be good to collect data on organic flocks in different countries on a routine basis to get more insight in the actual welfare status of organic broilers. E.g., on a theoretical basis it is suggested that organic conditions may best meet the behavioral needs of broiler chickens [3], but this might be highly dependent on the layout and the condition of the outdoor range, for example. Data collected here and in earlier studies [15, 16] indicate a considerable variation between flocks in contact dermatitis, due to climatic conditions affecting the litter conditions indoor and the condition of the outdoor range.

The data of broiler welfare on-farm of the three production systems (conventional, DRB, BLS) indicate that welfare performance is better in Better Life one star than in the other two systems, and that welfare performance of Dutch Retail Broiler flocks is better than conventional. This is in line with earlier studies including a smaller number of flocks [15, 16, 19] and also with a more theoretical comparison between the various production systems [3]. Most data were collected in 2017. Gait score is not assessed on a routine basis, but is considered to be one of the most important welfare issues in broiler chickens [20]; therefore trial data (2016-2019) from a project partner were included as these were the only recent data available. Gait scoring is time consuming, as currently automated systems are not available in practice, and therefore not routinely done. Development of new techniques such as video imaging offer opportunities for routinely collecting gait scores in the future (e.g., [21]). Also for behavior, data are lacking and at the moment, resource-based measures are used instead. It would be good to collect data on actual behaviour and use of enrichment instead of providing resource-based data, but this is also time consuming. Also here, future technological developments may provide opportunities for automated recording of broiler behavior [22]. Although on average DRB and BLS received better scores for the various indicators, this was not the case for scratches, which were more prevalent in BLS and DRB. This is likely to be caused by the higher activity of slower growing broiler chickens [23-25], increasing the risk for chickens walking over each other.

As far as we know, there is no literature on the overall welfare of broiler parent stock of fast and slower growing strains, especially also when the different slower growing breeds are compared. Although the table also includes estimations, it provides a good overview of the actual or estimated welfare status of parent stock of different broiler production systems. Restricted feeding, which is considered a welfare issue in broiler breeders [26], is applied at a much lower level or not at all in parent stock of slower growing broiler strains [27]. In addition to that, the table indicates that for several other welfare indicators parent stock of slower growing strains show more favorable scores. This is however not the case for all indicators: e.g., in the Netherlands mutilations are not allowed any more in female conventional breeders, whereas these are still allowed in parent stock of slower growing broiler strains.

For the hatchery phase, almost all data are lacking except these on the provision of early feeding. It is therefore not possible to compare the various production systems in this production phase.

Finally, for the end-of-life stage, there are also only few data available. More data, e.g. on fractures, were collected in other studies for fast growing flocks in Belgium [28] or a small number of fast and slower growing flocks in the Netherlands [19]. The Greenwell consortium however did not consider these data representative for the Dutch situation. This could have been caused by the fact that data on slower growing broiler flocks were only collected for a few flocks on one plant in Gerritzen et al. [19] or because the conditions in Belgium [28] are not comparable to the Dutch situation. The proportion of dead-on-arrival shows more favorable scores for DRB and BLS compared to conventional, whereas the proportion of breast, leg and wing bruises were more variable, e.g., BLS showing more wing bruises than conventional and DRB but DRB showing more breast bruises than conventional and BLS. It is highly recommended to collect more data on the end-of-life stage of broiler chickens, e.g. on the prevalence of fractures, trapped limbs and other injuries, as these are considered painful [28-30] and negatively affecting the welfare during this stage.

4.2 Part 2: overall welfare scores for broilers on-farm

In the present study we made the choice to assign welfare scores to individual welfare indicators according to the Welfare Quality approach [7], that was also applied to select the indicators in the first stage of development of the welfare model [6]. In literature, various methods have been used to assign a score to animal welfare within a sustainability model [31-33] and there is yet not a single generally applied method. In these papers, stakeholders or experts were consulted to assign scores to indicators which involves weighting and subjectivity. By applying the Welfare Quality model, we used expert opinion to assign scores to values for welfare indicators which also involves a degree of subjectivity. This should not be an issue as long as the method is transparent and sufficiently sensitive to variation in actual scores, and it is the only way to generate a total welfare score for a flock which is based on several indicators with a different scaling. However, it should be taken into account that the selection of experts determines the final outcome and that for example including other stakeholders may lead to a different scoring system [34]. A critical review of the Welfare Quality method to calculate flock scores led to the suggestion that a scoring system should at least (1) make sure that serious welfare problems are not overlooked or underestimated and (2) be transparent with respect to the ethical decisions that were taken [34]. We tried to meet these points in our current model, but we do not claim that our model is perfect; it is considered a way to include animal welfare in the overall sustainability assessment of broiler production systems, it includes the most important welfare issues in broiler chickens [20, 35], and it provides insight in differences between flocks with respect to the selected indicators. Furthermore, the Greenwell model is not considered static, but can be adjusted in case of new developments in the field [6].

In Welfare Quality, after generating scores for the individual indicators, an additional expert weighing was done to calculate criterion, principle and the total welfare score so that some indicators received higher weights, thus had a larger effect, on the total score. However, this resulted in a scoring system that appeared to be insensitive to the actual difference on the level of welfare indicators [36]. To prevent this, we decided to only include an expert weighing to generate scores from the various indicators on the same scale (0-100), and to give all indicators a similar weight to generate the total

welfare score (i.e., simple summing the individual scores). Inspection of the histogram with the flock scores per system showed a considerable variation between flocks within a system, and between the production systems, indicating that a simple summing of scores could be sufficiently sensitive to provide insight in the actual differences between flocks. However, assigning a different weight to indicators, e.g. to indicators that are perceived to have more serious effects on welfare, could have generated a different picture. Based on the existing literature on the welfare effects of the single indicators it is currently very difficult to assign different weights to the indicators and we therefore did not include this in the present model. This does however not mean that in a later stage, different weights can be assigned to the indicators.

One of the other drawbacks of the Greenwell model in the present study is that we did not have animal-based measures for all indicators, leading to absence of variation or little variation between flocks for the resource- or management-based indicators that have been included. Indeed, omitting these three resource-based indicators (early feeding, stocking density, environmental enrichment) showed that there was a relatively strong effect of these particular indicators on the distribution and value of the flock scores between production systems, although the overall difference between the production systems was not changed (from worst to best: conventional, DRB, BLS). Development of new methods, such as video imaging, can help to collect data on flock level that are thus far lacking and that can replace the resource-based measures (e.g., [22, 37]. An alternative for the current state-of-the art with respect to data collection could be not to include resource-based measures in the model. In that case no indication of behaviour will be included in the Greenwell model at all, which is also considered an unfavourable situation. Another drawback of the choice to simply sum the individual indicator scores to one Total Welfare Score is that compensation is allowed; i.e. a low score for a certain indicator can be compensated by a high score for another indicator, which was also one of the criticisms regarding Welfare Quality [34]. This can be overcome by e.g. reducing the total flock score when a certain threshold for an indicator is reached, as was done in [38] or by indicating minimum values that need to be obtained for each of the indicators. We did not apply this yet in the Greenwell model for broiler welfare on-farm but this might be a future option if we feel this would improve the model.

In the assessment of the different production systems with the Greenwell welfare model, most data for the on-farm broiler stage could be included because these were routinely collected by the processing plant or (temporarily) replaced by resource-based measures. It should however be noted that locomotion (gait score), one of the major welfare issues in broiler chickens [20], could not be included because of lack of data. Inspection of the limited flock data collected in small scale trials of a project partner indicated that these were in line with most other indicators with respect to production system differences. It is highly recommended to collect gait score data on a routine basis in commercial flocks and to include these data in the Greenwell welfare model.

As far as we know this is the first time that data on broiler welfare that have been collected for a large number of flocks from different production systems are published. The data and the welfare scores generated from these data show a better welfare level of BLS and DRB over conventional, confirming suggestions based on theoretical analyses [3] or small scale research [15, 19]. However, the data also illustrate that there is a large variation between flocks within the production systems and, when inspecting the total score for the animal-based measures, there is considerable overlap in welfare scores between production systems. This suggests that there is room for improvement in all production systems and provides options to use this Welfare assessment model to assist improvement of broiler welfare on individual farms by e.g. management advice. It is therefore advised to further study the flocks that overlap to find whether or not there is a consistent farm effect in e.g. the 10% best scores. This may lead to further management improvement and thus improvement of broiler welfare in commercial production systems. From previous research it is known that the individual broiler farm explains a large proportion of the variation in health and production indicators [39], which may suggest a management effect, and which may also be the case for broiler welfare outcomes.

Because most data could be collected for the broiler on-farm stage, the Greenwell model was only developed for broiler welfare on-farm. A similar methodology can be used to develop models for the broiler breeder, hatchery and end-of-life stage. Greenwell aims to include all stages of the broiler

production chain in the sustainability model. If and how the other stages of the broiler production chain can be included in the Greenwell model remains to be further decided.

4.3 Conclusions

In the Greenwell project, a sustainability assessment model will be developed for broiler production systems, including an environmental, an economic and an animal welfare assessment. Here we describe the development of the assessment model for broiler welfare on-farm, and tried to collect as much data possible on welfare indicators of four broiler production systems during the breeder, hatchery, on-farm and end-of-life stage.

The data collected and estimations over the year 2017 suggest a favourable welfare of broilers and parent stock for DRB and BLS compared to conventional, whereas for organic insufficient data could be collected to draw any conclusions about welfare in the different production phases. It is important to remind that most data were collected for broilers on-farm, whereas for the broiler breeder phase, end-of-life phase and hatchery phase the collected data were limited to very limited, respectively. Thus, there is a moderate to high degree of uncertainty for welfare in these phases for the broiler production systems that could be evaluated (conventional, Dutch Retail Broiler and Better Life one star).

The Greenwell welfare model provides a method to calculate an overall flock welfare score for broilers on-farm, based on routinely collected data in the broiler production chain or resource-based characteristics of production systems. This welfare model can be used in the overall sustainability assessment of broiler production systems. The welfare model shows, based on data of 2017 and 2018, a better welfare score of flocks in BLS over DRB and conventional production systems respectively . It also illustrates that scores within production systems show a high variability and that scores between production systems show considerable overlap, especially when the resource- and management based indicators are removed. This variation in welfare scores suggests room for improvement of broiler welfare within production systems and provides options to use this welfare assessment model to assist improvement of broiler welfare on individual farms by e.g. management advice.

Acknowledgements

The processing plant for broiler breeders is acknowledged for sharing their data with the consortium. Thea van Niekerk and the consortium members are acknowledged for reviewing the draft version of the report. The international broiler welfare experts that contributed to the score development are greatly acknowledged for their input.

References

1. Greenwell Wageningen: Wageningen Livestock research; 2018 [cited 2018 18-12-2018]. Available from: https://www.wur.nl/nl/Onderzoek-Resultaten/Onderzoeksinstituten/livestock-research/show-wlr/Greenwell-het-verenigen-van-dierenwelzijn-en-een-lage-milieubelasting-binnen-vleeskuiken-productiesystemen.htm.
2. Ellen HH, Leenstra F, R.A. VE, Groenestein K, Van Harn J, Van Horne P, et al. Broiler Production Systems in The Netherlands. Wageningen: Wageningen livestock Research, 2012 Contract No.: 619.
3. Stadig L. VLEESKUIKENCONCEPTEN IN NEDERLAND Een vergelijking op gebied van dierenwelzijn. Den Haag: Dierenbescherming, 2019.
4. Leinonen I, Williams AG, Wiseman J, Guy J, Kyriazakis I. Predicting the environmental impacts of chicken systems in the United Kingdom through a life cycle assessment: Broiler production systems. Poultry Science. 2012;91(1):8-25. doi: 10.3382/ps.2011-01634.
5. Mulder M, Zomer S, Benning T, Leenheer J. Economische effecten van 'Kip van Morgen' Kosten en baten voor consumenten van een collectieve afspraak in de pluimveehouderij. Autoriteit Consument en Markt, 2014.
6. de Jong IC. Development of the 'animal welfare' dimension within the Greenwell sustainability assessment model: 1. justification of the selection of indicators. Wageningen: Wageningen Livestock Research, 2019.
7. Welfare Quality. The Welfare Quality Assessment Protocol for Broiler Chickens and Laying Hens. Lelystad: The Welfare Quality Consortium; 2009.
8. de Jong IC, van Harn J, Gunnink H, Hindle VA, Lourens A. Footpad dermatitis in Dutch broiler flocks: Prevalence and factors of influence. Poultry Science. 2012;91(7):1569-74. doi: 10.3382/ps.2012-02156.
9. Pluimned. IKB Kip version 5: Pluimned; 2019 [8-11-2018]. Available from: https://pluimned.avined.nl/sites/pluimned.avined.nl/files/8-beoordelingssysteem_vleeskuikens_ikb_kip_-_versie_5_-_190601.pdf.
10. Stichting Diergeneesmiddelen Autoriteit. Het gebruik van antibiotica bij landbouwhuisdieren in 2017. Utrecht: 2017.
11. Saatkamp HW, Vissers LSM, Van Horne P, IC dJ. Transition from Conventional Broiler Meat to Meat from Production Concepts with Higher Animal Welfare: Experiences from The Netherlands. Animals. 2019;9(8):483-. doi: https://doi.org/10.3390/ani9080483
12. Avined. ANTIBIOTICUMGEBRUIK PLUIMVEESECTOR IN 2018 en de trends van afgelopen jaren 2019 [31-10-2019]. Available from: http://www.avined.nl/sites/www.avined.nl/files/antibioticagebruik_-_sectorrapportage_2018.pdf.
13. de Jong IC, Hindle VA, Butterworth A, Engel B, Ferrari P, Gunnink H, et al. Simplifying the Welfare Quality((R)) assessment protocol for broiler chicken welfare. Animal. 2016;10(1):117-27. doi: 10.1017/s1751731115001706.
14. Blokhuis H, Miele M, Veissier I, Jones B. Improving farm animal welfare science and society working together: The welfare quality approach2013. 1-232 p.
15. De Jong IC, Gunnink H, Hindle VA. Implementation of the Welfare Quality®broiler assessment protocol – final report. Wageningen: Wageningen Livestock Research, 2015 Rapport 833.
16. De Jong IC, Perez Moya T, Gunnink H, Van den Heuvel H, Hindle VA, Mul M, et al. Simplifying the Welfare Quality assessment protocol for broilers. Lelystad: Wageningen Livestock Research, 2011 Rapport 533.
17. Souillard R, Reperant JM, Experton C, Huneau-Salaun A, Coton J, Balaine L, et al. Husbandry Practices, Health, and Welfare Status of Organic Broilers in France. Animals. 2019;9(3). doi: 10.3390/ani9030097. PubMed PMID: WOS:000464323300003.
18. Wilhelmsson S, Yngvesson J, Jonsson L, Gunnarsson S, Wallenbeck A. Welfare Quality (R) assessment of a fast-growing and a slower-growing broiler hybrid, reared until 10 weeks and fed a low-protein, high-protein or mussel-meal diet. Livestock Science. 2019;219:71-9. doi: 10.1016/j.livsci.2018.11.010.
19. Gerritzen M, Verkaik J, Reimert H, Gunnink H, Van Hattum T, de Jong IC. Letsel en schade bij vleeskuikens als gevolg van vangen, transport en handelingen aan de slachtlijn Wageningen: Wageningen Livestock Research, 2019 1107.
20. EFSA. Scientific Opinion on the influence of genetic parameters on the welfare and the resistance to stress of commercial broilers. EFSA journal. 2010; 8(7):1666 -.

21. Aydin A, Bahr C, Berckmans D. Automatic classification of measures of lying to assess the lameness of broilers. Anim Welf. 2015;24(3):335-43. doi: 10.7120/09627286.24.3.335.
22. Dawkins MS, Cain R, Merelie K, Roberts SJ. In search of the behavioural correlates of optical flow patterns in the automated assessment of broiler chicken welfare. Applied Animal Behaviour Science. 2013;145(1-2):44-50. doi: 10.1016/j.applanim.2013.02.001.
23. Rayner A, Newberry RC, Vas J, Butterworth A, Mullan S. Growing slowly with more space: effects on "positive behaviours" in broiler chickens. Abstracts of the ISAE; Bergen 2019.
24. Torrey S, Liu Z, Caston L, Nascimento dos Santos M, Rothschild D, Kiarie E, et al. Differences in behavioural time budget between conventional and slow growing broiler chickens. Poultry Science Association Meeting; Montreal: Poultry Science Association; 2019.
25. Bokkers EAM, Koene P. Behaviour of fast- and slow growing broilers to 12 weeks of age and the physical consequences. Applied Animal Behaviour Science. 2003;81(1):59-72. doi: 10.1016/s0168-1591(02)00251-4.
26. De Jong IC, Guemene D. Major welfare issues in broiler breeders. Worlds Poult Sci J. 2011;67(1):73-81. doi: 10.1017/s0043933911000067.
27. Decuypere E, Hocking PM, Tona K, Onagbesan O, Bruggeman V, Jones EKM, et al. Broiler breeder paradox: a project report. Worlds Poult Sci J. 2006;62(3):443-53. doi: 10.1079/wps2005107.
28. Jacobs L, Delezie E, Duchateau L, Goethals K, Tuyttens FAM. Impact of the separate pre-slaughter stages on broiler chicken welfare. Poultry Science. 2017;96(2):266-73. doi: 10.3382/ps/pew361.
29. Kittelsen KE, Granquist EG, Vasdal G, Tolo E, Moe RO. Effects of catching and transportation versus pre-slaughter handling at the abattoir on the prevalence of wing fractures in broilers. Anim Welf. 2015;24(4):387-9. doi: 10.7120/09627286.24.4.387.
30. EFSA committee on Animal Health and Animal Welfare. Scientific opinion on slaughter of animals: poultry. EFSA journal. 2019;17(11):5489. doi: https://doi.org/10.2903/j.efsa.2019.5849.
31. van Asselt ED, van Bussel LGJ, van Horne P, van der Voet H, van der Heijden GWAM, van der Fels-Klerx HJ. Assessing the sustainability of egg production systems in The Netherlands. Poultry Science. 2015;94(8):1742-50. doi: 10.3382/ps/pev165.
32. Tallentire CW, Edwards SA, Van Limbergen T, Kyriazakis I. The challenge of incorporating animal welfare in a social life cycle assessment model of European chicken production. The International Journal of Life Cycle Assessment
2018. doi: https://doi.org/10.1007/s11367-018-1565-2.
33. Castellini C, Boggia A, Cortina C, Dal Bosco A, Paolotti L, Novelli E, et al. A multicriteria approach for measuring the sustainability of different poultry production systems. Journal of Cleaner Production. 2012;37:192-201. doi: 10.1016/j.jclepro.2012.07.006.
34. Sandoe P, Corr SA, Lund TB, Forkman B. Aggregating animal welfare indicators: can it be done in a transparent and ethically robust way? Anim Welf. 2019;28(1):67-76. doi: 10.7120/09627286.28.1.067.
35. EFSA. Scientific Opinion on the use of animal-based measures to assess welfare of broilers. EFSA Journal. 2012;10(7):2774. doi: doi:10.2903/j.efsa.2012.2774.
36. Buijs S, Ampe B, Tuyttens FAM. Sensitivity of the Welfare Quality® broiler chicken protocol to differences between intensively reared indoor flocks: which factors explain overall classification? Animal. 2017;11(2):244-53.
37. Aydin A, Pluk A, Leroy T, Berckmans D, Bahr C. AUTOMATIC IDENTIFICATION OF ACTIVITY AND SPATIAL USE OF BROILER CHICKENS WITH DIFFERENT GAIT SCORES. Transactions of the Asabe. 2013;56(3):1123-32.
38. Jacobs L, Delezie E, Goethals K, Ampe B, Duchateau L, Tuyttens FAM. Vleeskippenwelzijn tijdens de pre-slachtfase Evaluatieprotocol en Online integratie-tool. Melle, Belgium: ILVO, 2017.
39. de Jong IC, Riel JW van. Relative contribution of production chain phases to health and performance of broiler chickens: a field study. Poultry Science. 2020;99:179-88.

Appendix 1 Graphic representation of indices and scores according to Welfare Quality

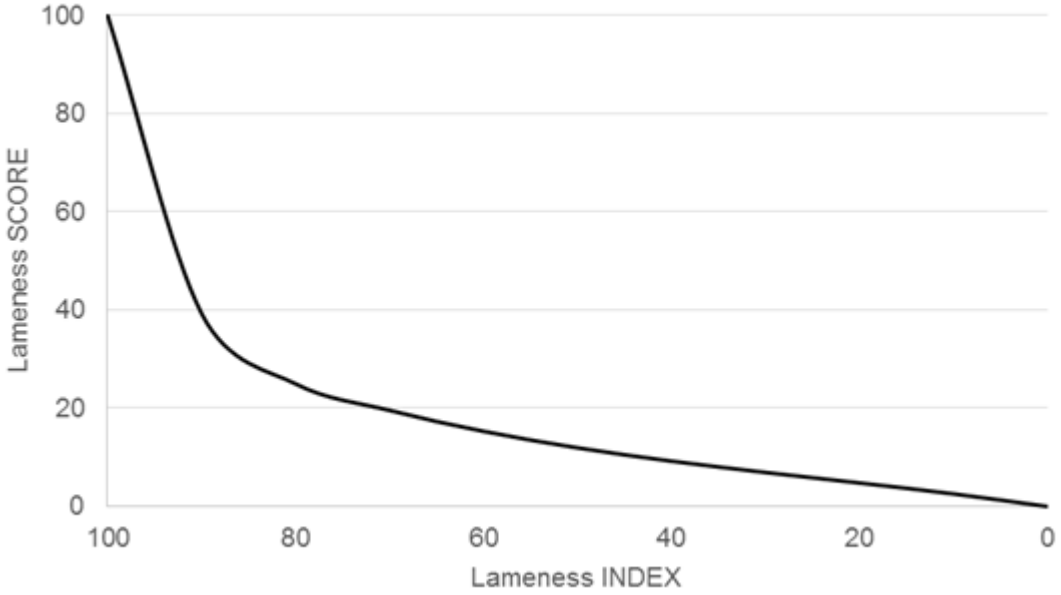


Figure 1.1 *Spline function representing the relationship between the lameness index and the lameness score (source: [7]).*

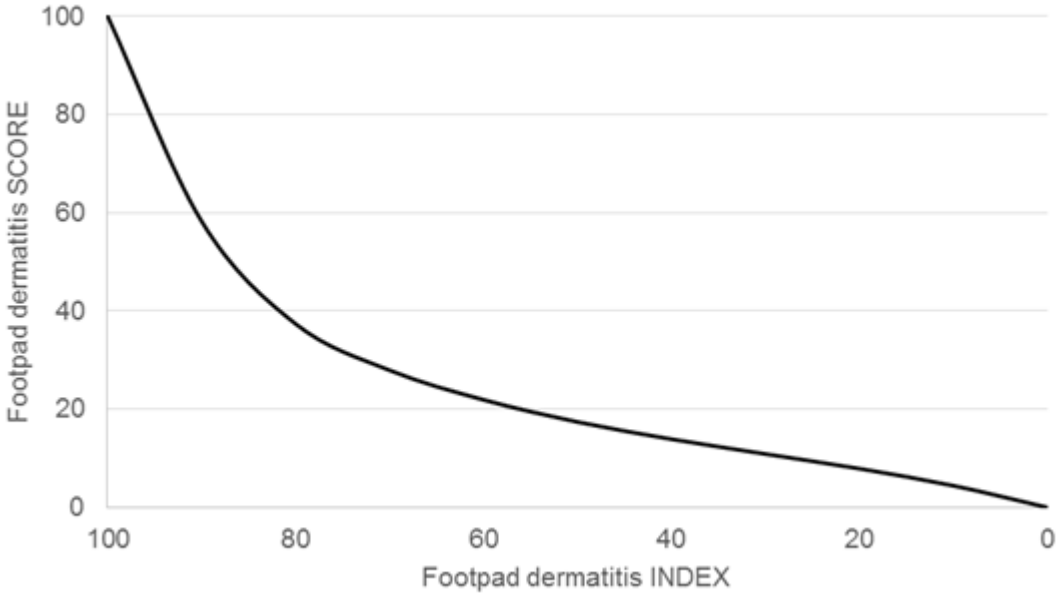


Figure 1.2 *Spline function representing the relationship between the footpad dermatitis index and the footpad dermatitis score (source: [7]).*

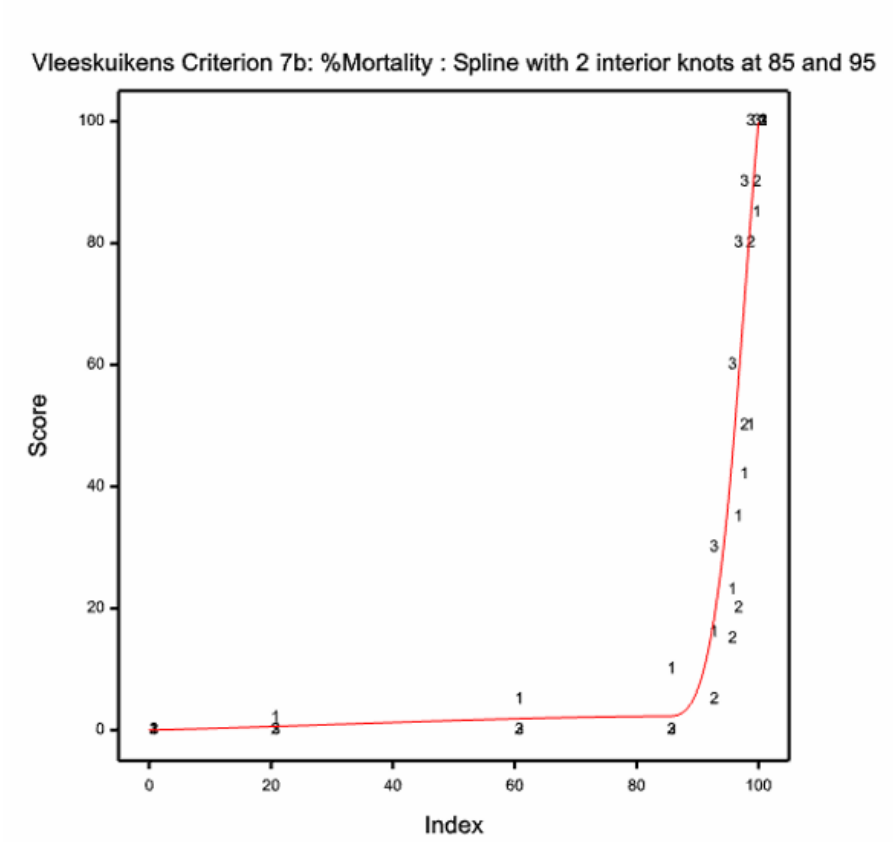


Figure 1.3 *Spline function representing the relationship between the mortality index and the mortality score (source:[16]).*

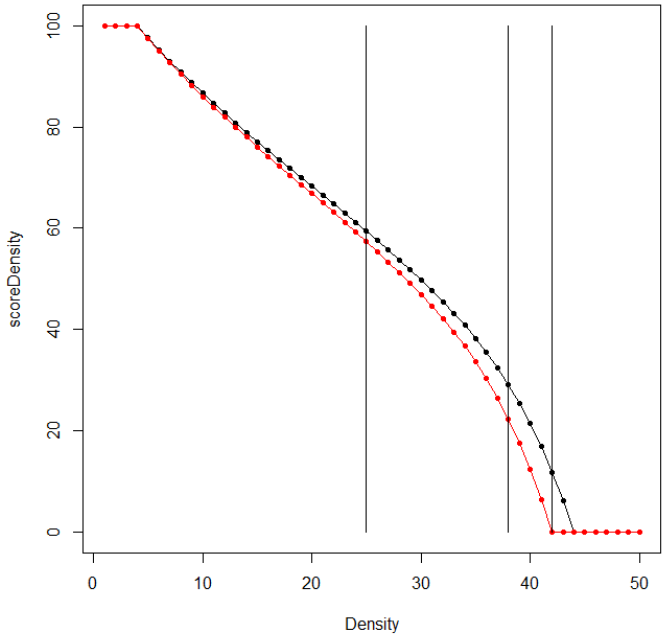


Figure 1.4 *Spline function representing the relationship between the stocking density and the score. The black line is the original spline function according to [7], red represents the spline function which is applied in the current report and where the maximum stocking density has been adjusted to 42 kg/m².*

Appendix 2 Expert consultation

Below we include the document that has been sent to the experts that have been consulted to determine the new spline functions to calculate scores for broiler welfare on-farm. In addition, two indicators related to broiler welfare during the end-of-life stage have been included in the expert consultation, so that these can easily be applied in a later stage.

Additional expert consultation for construction of indicator scores for broiler welfare on-farm or during the end-of-life stage, according to the Welfare Quality method

Ingrid de Jong

October, 2019

Wageningen Livestock Research

1. Introduction

Wageningen Livestock Research, together with four partners in the broiler production chain and the Ministry of Agriculture, Nature and Food Quality, is currently developing a sustainability assessment model for broiler chicken production within the 'Greenwell' project. This sustainability assessment model consists of three sub-models, i.e. a welfare assessment model, an environmental impact model, and an economical model.

Regarding the welfare assessment model, the Greenwell project chose to base the model on the existing Welfare Quality® assessment protocol for broiler chickens (on-farm stage)[1]. Regarding the end-of-life stage, the hatchery stage and the broiler breeder stage, new models have been developed, based on the welfare criteria and measures as defined by Welfare Quality®, but with specific measures regarding the stage of the production chain. If present, existing models, such as WellTrans[2] for the end-of-life stage, were taken into account[1].

A few new animal-based indicators were considered by the Greenwell consortium for the assessment of broiler welfare on-farm and are included in final selection of key-indicators. Reasons for adding new indicators in addition to indicators of existing models is that in the existing Welfare Quality® protocol, indicators for some criteria are lacking (e.g., for social behaviour), or are subject to discussion between scientists because of lack of validity (e.g., Qualitative Behaviour Assessment and Touch Test in the Welfare Quality® broiler assessment protocol[3]), and that new indicators of animal welfare are still being developed.

Further, as Greenwell will use data collected in the broiler chain, there are sometimes differences in the scoring method as compared to Welfare Quality®, which requires that the current function to calculate an indicator score, as described by Welfare Quality, should be adjusted. An example is the scoring for hock burn, which has five classes according to Welfare Quality®, but only two classes if done by the plant, due to the high slaughter line speed.

Indicators requiring new calculations for indicator scores

The following indicators for broiler welfare on-farm and during the end-of-life stage need adjustment of the spline function to calculate the indicator score, or a spline function needs to be developed because the indicator is new (Table 1) :

[1] De Jong, I.C., 2019. Development of the 'animal welfare' dimension within the Greenwell sustainability assessment model: 1. justification of the selection of indicators. Wageningen Livestock Research Report 1194.

[2] JACOBS, L., DELEZIE , E., GOETHALS, K., AMPE, B., DUCHATEAU, L. & TUYTTENS, F. A. M. 2017. Vleeskippenwelzijn tijdens de pre-slachtfase Evaluatieprotocol en Online integratie-tool. Melle, Belgium: ILVO.

[3] The relationship between fear of human and lameness in broilers (2018). Vasdal, G., de Jong, I., Moe, R. O., Granquist, E.G. Animal, 12, 334-33910.1017/s1751731119000466

Table 1 : list of indicators for which a new calculation is required in Greenwell, with reason.

Indicator	**Reason**
Early feeding (post-hatch)	New indicator, relates to absence of hunger and thirst
Hock burn	Scoring method at the plant differs from Welfare Quality®
Breast burn/irritation	Replaces breast blisters, as these are hardly observed in broiler chickens; spline calculation in the Welfare Quality® protocol needs to be adjusted.
Scratches	New indicator, relates to absence of injuries
Presence of enrichment/outdoor range/natural light	New indicator ; this resource-based measure will temporarily replace the animal-based measure 'species specific behaviour' until these data can/will be collected on-farm
Bruises	Indicator for welfare during end-of-life stage, no calculation developed by Welfare Quality®
Dead-on-arrival	Indicator for welfare during end-of-life stage, no calculation developed by Welfare Quality®

Methodology

Because we will use the Welfare Quality® calculation of scores for the indicators, we would like to follow the same methodology for the (additional) indicators in the Greenwell welfare model for broilers on-farm. This enables us to calculate a final welfare score for broiler flocks of different production systems. Please note that we will only calculate scores per indicator, and that at this moment we will not generate criterion and principle scores as is done in the Welfare Quality® assessment protocol.

Expert consultation

We would like to ask your help to define functions to calculate scores for indicators that have been added in the Greenwell welfare assessment model. An appropriate method would be to generate spline functions for the indicators % chickens with scratches, hock burn, breast burn/irritation, bruises, and % dead-on-arrival. A decision tree will be applied to calculate a score for the indicators 'early feeding' and 'enrichment/natural light/veranda/outdoor range', according to the approach of Welfare Quality®. In order to do this, we need an additional expert opinion on each of these indicators.

Instructions to fill in the tables

In the current document, you will find tables with virtual data with regard to the total prevalence of birds with injuries, hock burn, breast burn/irritation, bruises and the prevalence of dead-on-arrival. Further, for your information, a histogram is presented showing the distribution of the indicator prevalence for a one-year period as collected at slaughter. This includes different farm types, i.e., conventional farms with fast growing strains and farms with slower growing broilers, lower stocking densities and environmental enrichment. You are kindly asked to keep this information confidential. Further, you will find two decision trees regarding the presence or absence of early feeding and environmental enrichment/outdoor range.

For each indicator I would like to ask you to give a score for the level of welfare for each of the virtual farms (see tables below). The welfare score is always between 0 and 100; a score of 0 refers to the lowest level of welfare, and a score of 100 refers to the highest possible level of welfare. As a point of reference, you are advised to keep in mind that a welfare score below 20 refers to an unacceptable situation, a score between 20 and 50 would be just acceptable, a score between 50 and 80 would refer to an enhanced welfare situation, and in case of a score > 80, the welfare situation would be considered excellent.

Similarly, I would like to ask you to assign a score for either or not using early feeding, and the different combinations of natural light/enrichment/outdoor range.

If you feel that you do not have the expertise to fill in one of the tables, please indicate this in your email and leave it empty.

Your names will not be disclosed, we will just mention the number of experts that contributed to this consultation.

Thank you very much for your collaboration. If necessary, you can always contact me at ingrid.dejong@wur.nl, +31 317 480 589.

2. Hock burn

Please assign a score to the prevalence of hock burn on the 14 virtual farms as shown in Table 2. Please keep in mind the meaning of the scores as indicated below the table. As a reference, a histogram of the prevalence of hock burn during a one year period for different types of broiler farms (conventional to higher-welfare systems) is presented below the table.

Hock burn is scored as follows[1]:
Score 0: no evidence of hock burn, or a discoloration of a size of 0.5 cm^2 at maximum
Score 1: evidence of hock burn, any brown or black discoloration of the hock of at least 0.5 cm^2 in size

Table 2: virtual dataset and scoring table for hock burn.

Farm	Prevalence of hock burn	Score (100=perfect)
1	0	100
2	1	90
3	2	80
4	3	70
5	4	60
6	6	50
7	8	40
8	11	30
9	15	20
10	21	10
11	30	0
12	50	0
13	70	0
14	100	0

Meaning of scores:									
0	unacceptable	20	acceptable	55	enhanced	80	excellent	100	

[1] https://pluimned.avined.nl/sites/pluimned/files/na_1_mrt_17_8-beoordelingssysteem_vleeskuikens_ikb_kip_-_versie_4_-_170301.pdf

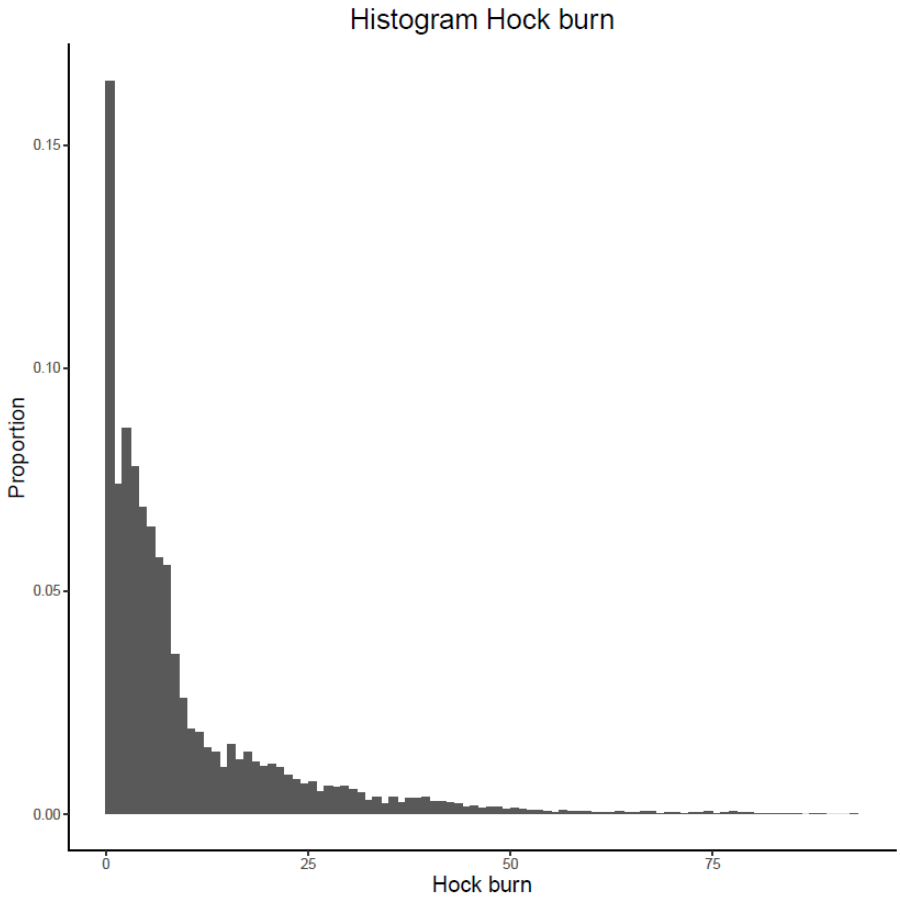


Figure 1: histogram of prevalence of hock burn (score 1) during 1 year, in different types of broiler systems.

3. Breast burn/irritation

Please assign a score to the prevalence of breast burn/irritation on the 14 virtual farms as shown in Table 3. Please keep in mind the meaning of the scores as indicated below the table. As a reference, a histogram of the prevalence of breast burn during a one year period for different types of broiler farms (conventional to higher-welfare systems) is presented below the table.

Breast burn is scored as follows[6]:
Score 0: no evidence of breast burn/irritation or discoloured area/lesion smaller than 0.5 cm^2
Score 1: A brown/black discoloured area or lesion larger than 0.5 cm^2 on the breast

[6] https://pluimned.avined.nl/sites/pluimned/files/na_1_mrt_17_8-beoordelingssysteem_vleeskuikens_ikb_kip_-_versie_4_-_170301.pdf

Table 3: virtual dataset and scoring table for breast burn/irritation.

Farm	Prevalence of breast burn/irritation	Score (100=perfect)
1	0	100
2	0.1	90
3	0.2	80
4	0.3	75
5	0.4	70
6	0.6	65
7	0.9	60
8	1.5	50
9	3.0	40
10	5.0	20
11	9.0	0
12	15.0	0
13	25.0	0
14	40.0	0

Meaning of scores:									
0	unacceptable	20	acceptable	55	enhanced	80	excellent	100	

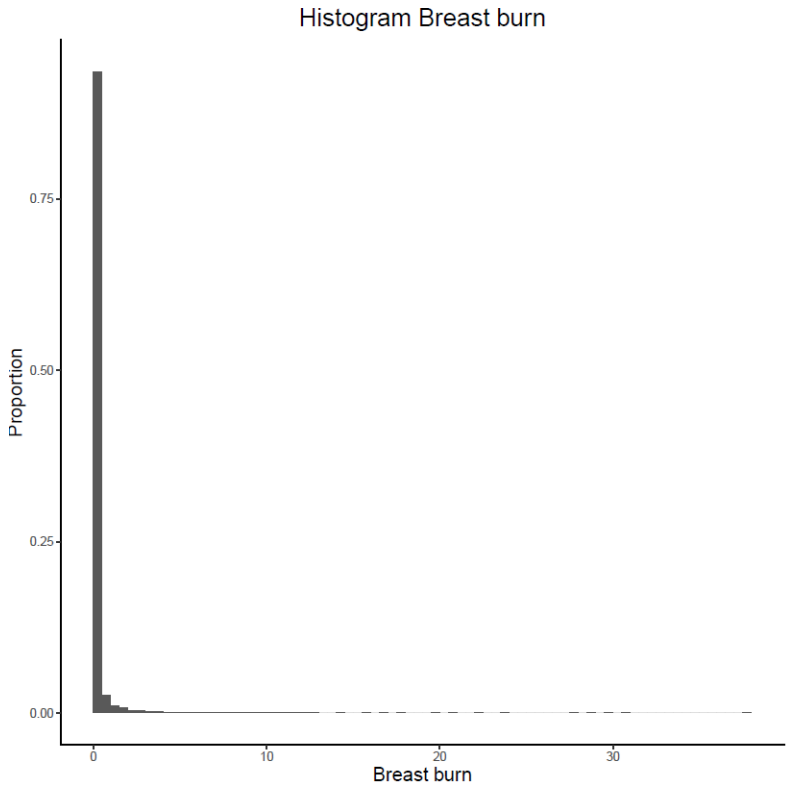


Figure 2: histogram of prevalence of breast burn (score 1) during 1 year, in different types of broiler systems.

4. Scratches

Please assign a score to the prevalence of scratches on the 13 virtual farms as shown in Table 4. Please keep in mind the meaning of the scores as indicated below the table. As a reference, a histogram of the prevalence of scratches during a one year period for different types of broiler farms (conventional to higher-welfare systems) is presented below the table.

Scratches are scored as follows[7]:
Score 0: no evidence of scratches, or less or smaller than defined for score 1 on the breast or thigh area
Score 1: a score of 1 was assigned when 3 scratches > 2cm were observed (fresh, or scab or crust) or when a wound (open skin, either or not covered with a crust) was observed on the breast or thigh area

[7] https://pluimned.avined.nl/sites/pluimned/files/na_1_mrt_17_8-beoordelingssysteem_vleeskuikens_ikb_kip_-_versie_4_-_170301.pdf

Table 4: virtual dataset and scoring table for scratches.

Farm	Prevalence of scratches	Score (100=perfect)
1	0	100
2	0.4	90
3	0.8	80
4	1.2	70
5	1.6	55
6	2.0	40
7	2.8	20
8	4.0	10
9	6.0	0
10	8.0	0
11	11.0	0
12	15.0	0
13	20.0	0

Meaning of scores:									
0	unacceptable	20	acceptable	55	enhanced	80	excellent	100	

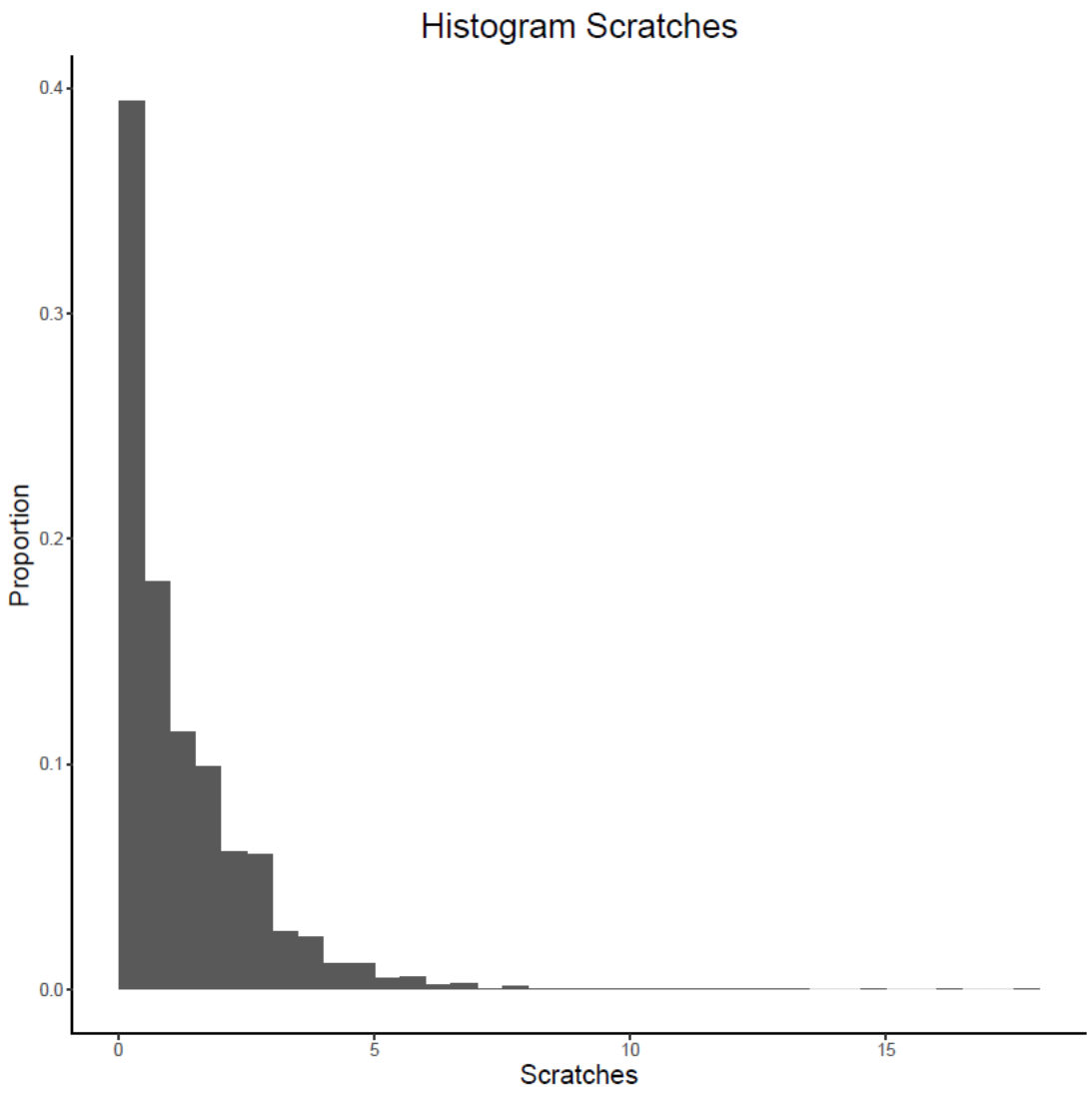


Figure 3: histogram of prevalence of scratches (score 1) during 1 year, in different types of broiler systems.

5. Dead-on-arrival

Please assign a score to the prevalence of dead-on-arrival on the 14 virtual farms as shown in Table 5. Please keep in mind the meaning of the scores as indicated below the table. As a reference, a histogram of the prevalence of dead-on-arrival during a one year period for different types of broiler farms (conventional to higher-welfare systems) is presented below the table. Dead-on-arrival are all birds that are found dead upon arrival at the slaughter plant.

Table 5: virtual dataset and scoring table for dead-on-arrival.

Farm	Prevalence of dead-on-arrival	Score (100=perfect)
1	0	100
2	0.01	90
3	0.02	80
4	0.03	70
5	0.04	60
6	0.06	55
7	0.09	40
8	0.14	30
9	0.20	20
10	0.50	0
11	1.00	0
12	2.00	0
13	3.00	0
14	5.00	0

Meaning of scores:									
0	unacceptable	20	acceptable	55	enhanced	80	excellent	100	

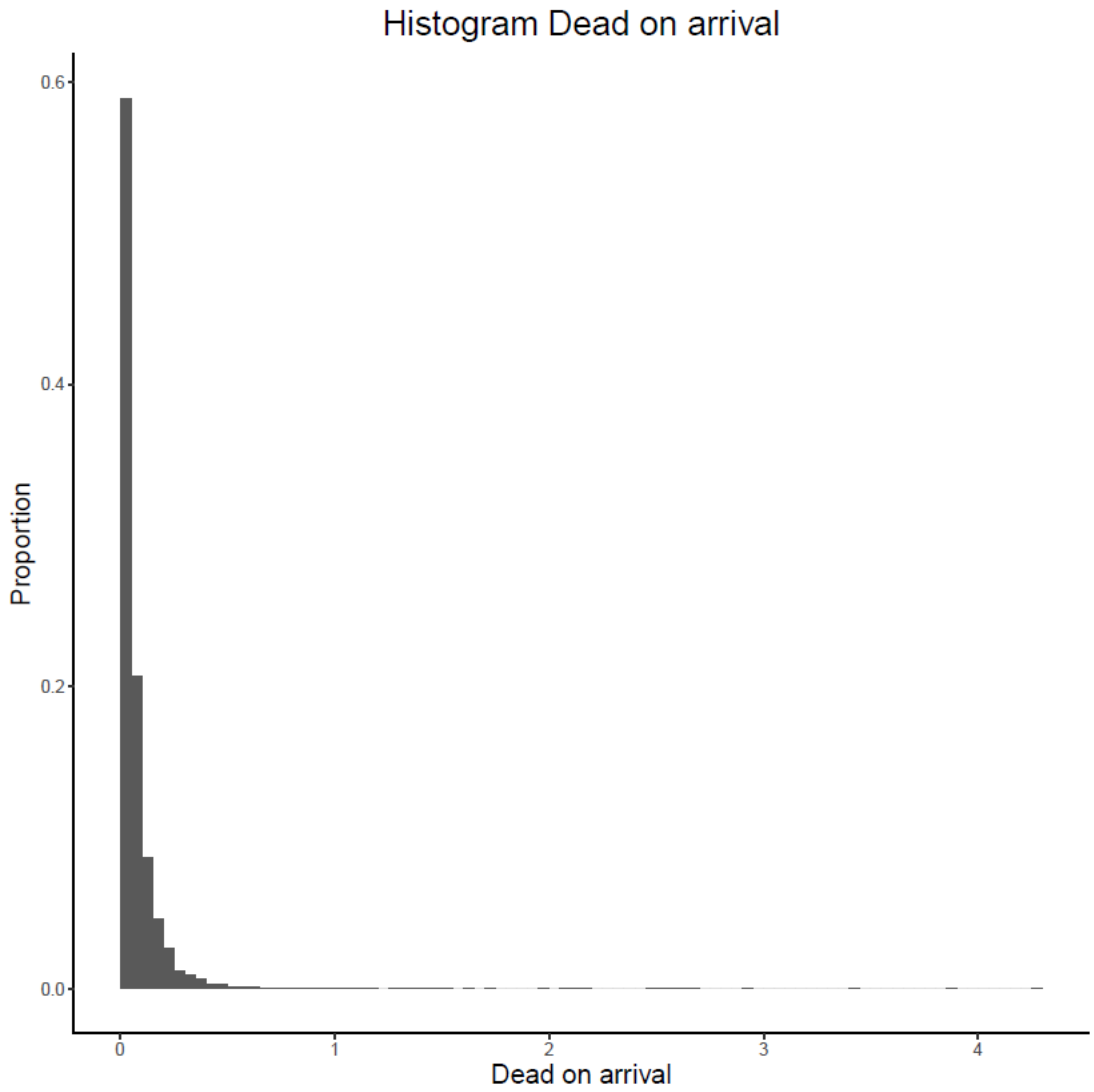


Figure 4: histogram of prevalence of dead-on-arrival (score 1) during 1 year, in different types of broiler systems.

6. Bruises

Please assign a score to the prevalence of bruises on the 14 virtual farms as shown in Table 6. Please keep in mind the meaning of the scores as indicated below the table. As a reference, a histogram of the prevalence of bruises during a one year period for different types of broiler farms (conventional to higher-welfare systems) is presented below the table.

Bruises are scored as follows[8]:
Score 0: no bruises on breast, legs or wings or bruises smaller than 1 (legs, breast) or 2 cm^2 (wings)
Score 1: One or more bruises present on the breast, legs or wings. On breast and legs, these need to be larger than 1 cm^2. On the wings, these need to be larger than 2 cm^2. For the wings, bruises on the tip of the wing are excluded. Bruises might, but should not necessarily, be accompanied by fractures.

Bruises on breast, legs and wings are scored separately. We will here use the sum of the proportions of chickens with breast, legs and wing bruises. Data analysis showed that correlations were low for the different types of bruises, indicating that flocks with e.g. wing bruises not necessarily also have breast or leg bruises.

[8] https://pluimned.avined.nl/sites/pluimned/files/na_1_mrt_17_8-beoordelingssysteem_vleeskuikens_ikb_kip_-_versie_4_-_170301.pdf

Table 6: virtual dataset and scoring table for the summed proportion of chickens with leg, breast and wing bruises.

Farm	Summed prevalence of leg, breast and wing bruises	Score (100=perfect)
1	0	100
2	1	80
3	2	70
4	3	55
5	4	45
6	5	40
7	6	35
8	7	30
9	8	25
10	9	20
11	11	0
12	15	0
13	22	0
14	30	0

Meaning of scores:									
0	unacceptable	20	acceptable	55	enhanced	80	excellent	100	

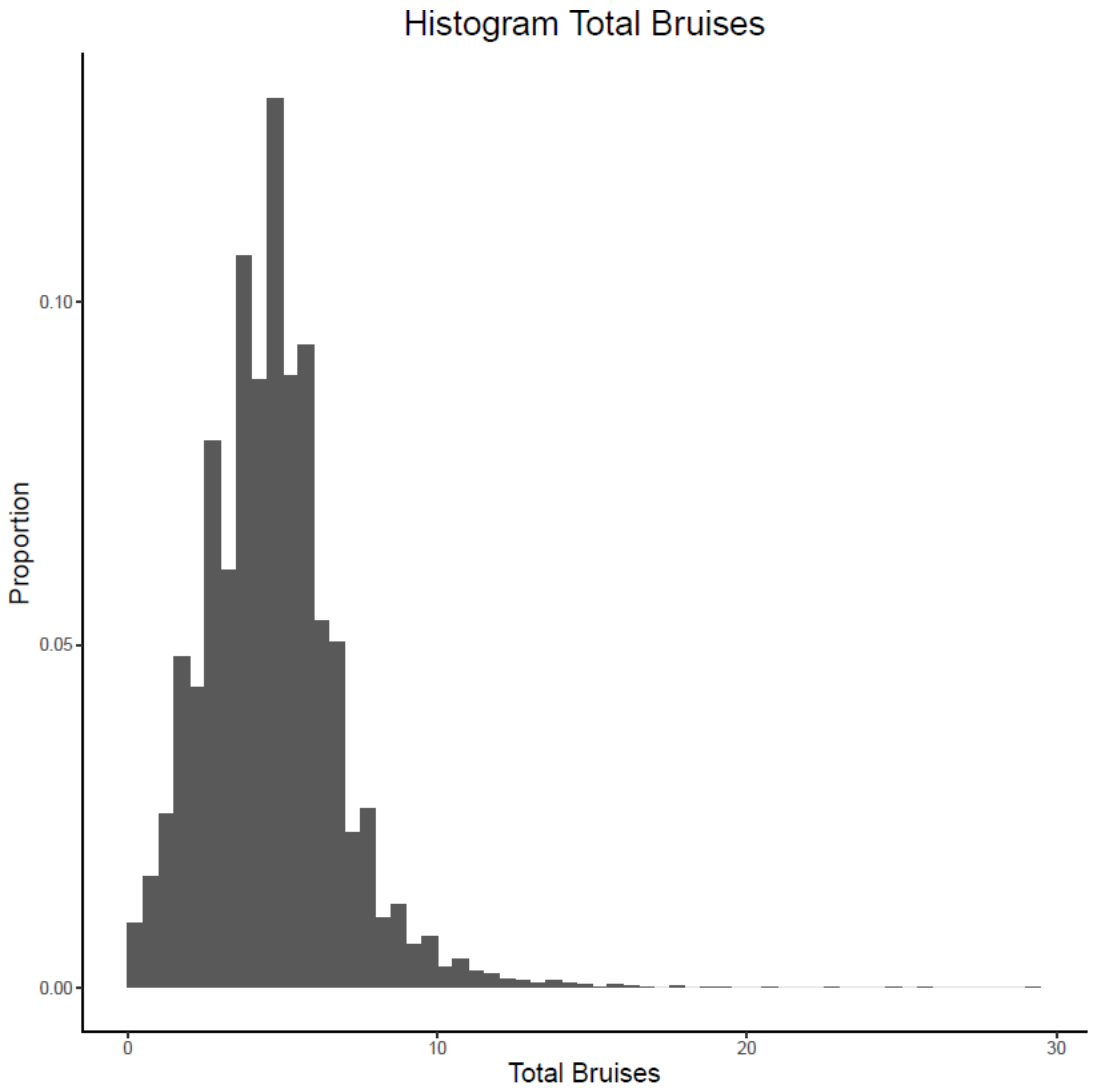


Figure 5: histogram of prevalence of the sum of chickens with leg, breast and wing bruises (score 1) during 1 year, in different types of broiler systems

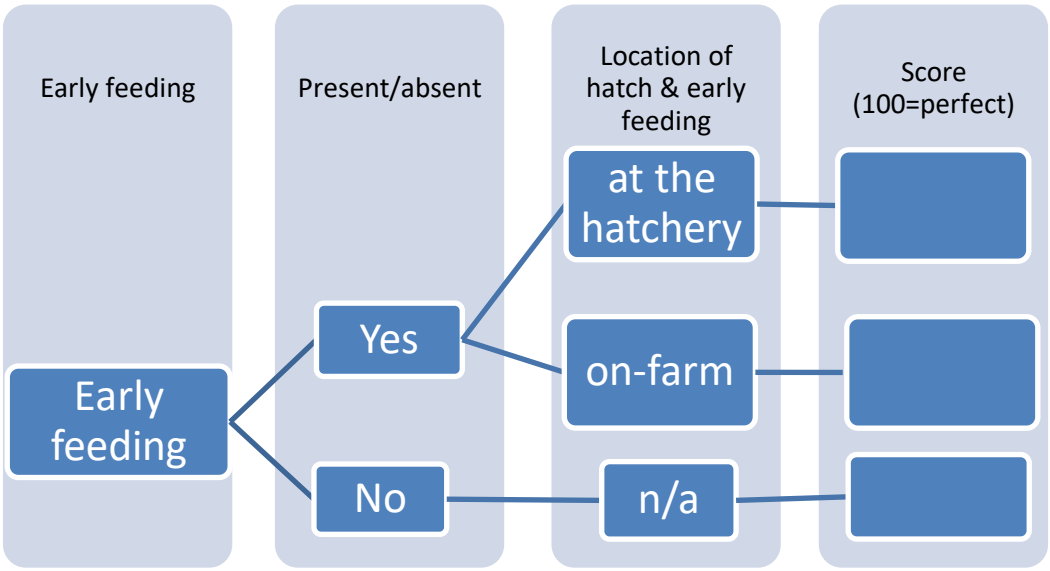
7. Early feeding

It has been shown that feed deprivation of 36 hours or more after hatching increases the risk for mortality in later life. Although the yolk sac provides energy during the first hours post-hatch, newly hatched chicks may suffer from hunger and thirst as it may take long before they receive their first feed and water, depending on e.g. hatching moment (early or late hatchers) and transport time [9]. To prevent hunger and thirst due to feed and water deprivation post-hatch, broiler can be fed in the hatchery and transported to the farm afterwards, or can hatch in the broiler house (in that case 18-days incubated eggs are placed in the broiler house). Currently, less than 10% of the broilers hatch in the house in the Netherlands and a small proportion of chickens receives feed in the hatchery.

Please assign a score to the different early feeding systems post-hatch as indicated in the decision tree.

[9] DE JONG, I. C., VAN RIEL, J., BRACKE, M. B. M. & VAN DEN BRAND, H. 2017. A 'meta-analysis' of effects of post-hatch food and water deprivation on development, performance and welfare of chickens. *PLoS One,* 12**,** e0189350.

Decision tree 1. Different systems for early feeding (click on the cell to fill in the score).



Meaning of scores:									
0	unacceptable	20	acceptable	55	enhanced	80	excellent	100	

8. Natural light and environmental enrichment

In the ideal situation, the behaviour of broiler chickens scored and included in the welfare assessment. However, there are practical limitations to collect these data in practice, although in the future this might be possible due to development of e.g. sensor techniques. Therefore, we included the presence or absence of resources such as natural light, environmental enrichment and (outdoor) ranges in the assessment model, until we would be available to record broiler behaviour at a large scale.

Currently in commercial systems in the Netherlands, there is variation in the application of natural light, the presence of a range or veranda, and the number of environmental enrichments that is provided. The latter can be of different types, such as perches, platforms, pecking objects (pecking stones) and bales. Sometimes several enrichment types are present, e.g. elevated resting place and pecking objects. We ask you to assign a score to the different combinations of natural light, range and environmental enrichment as indicated below. For clarity, the decision tree is presented as a table.

Natural light inside the house	Veranda/outdoor	Number of additional enrichment types in the house	Score (100=perfect)
No	No	0	
		1	
		2	
		≥ 3	
	Covered veranda	0	
		1	
		2	
		≥ 3	
	Outdoor range	0	
		1	
		2	
		≥ 3	
	Covered veranda AND outdoor range	0	
		1	
		2	
		≥ 3	
Yes	No	0	
		1	
		2	
		≥ 3	
	Covered veranda	0	
		1	
		2	
		≥ 3	
	Outdoor range	0	
		1	
		2	
		≥ 3	
	Covered veranda AND outdoor range	0	
		1	
		2	
		≥ 3	

Meaning of scores:

0 unacceptable	20 acceptable	55 enhanced	80 excellent 100

Appendix 3 List of experts that have been consulted

For the expert consultation as indicated in Appendix 1, the following experts have been consulted:

(1) Researcher broiler welfare, Welfare Quality expert, Italy
(2) Researcher broiler welfare, Welfare Quality expert, Norway[1]
(3) Researcher broiler welfare with knowledge of Welfare Quality measures, Germany[1]
(4) Researcher broiler welfare, with knowledge of Welfare Quality measures, USA
(5) Researcher broiler welfare, Welfare Quality expert, UK
(6) Researcher poultry welfare, Welfare Quality expert, The Netherlands
(7) Researcher broiler welfare, Welfare Quality expert, The Netherlands
(8) Researcher broiler and pig welfare, Welfare Quality expert, Spain
(9) Expert on welfare during the end-of-life stage, The Netherlands[2]

There were no experts that did not want to participate in the expert consultation.

[1] These experts indicated that they were unable to assign scores to 'early feeding'
[2] This expert only contributed with respect to the indicators for the end-of-life stage

Appendix 4 Graphical representation of new spline functions to calculate indicator scores

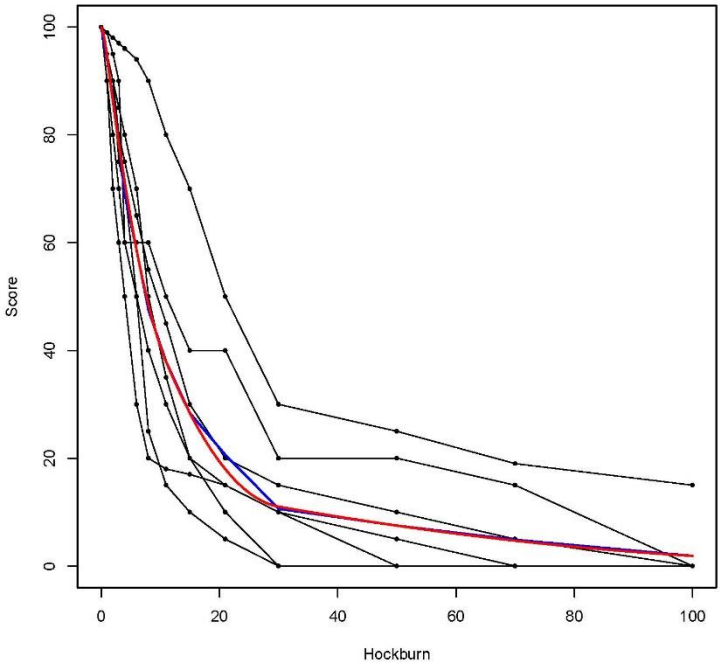


Figure 4.1 *Spline function representing the relationship between the prevalence of hock burn and the score. The black lines are the scores of the individual experts; the blue line represents the average; the red line represents the spline function derived of the individual scores.*

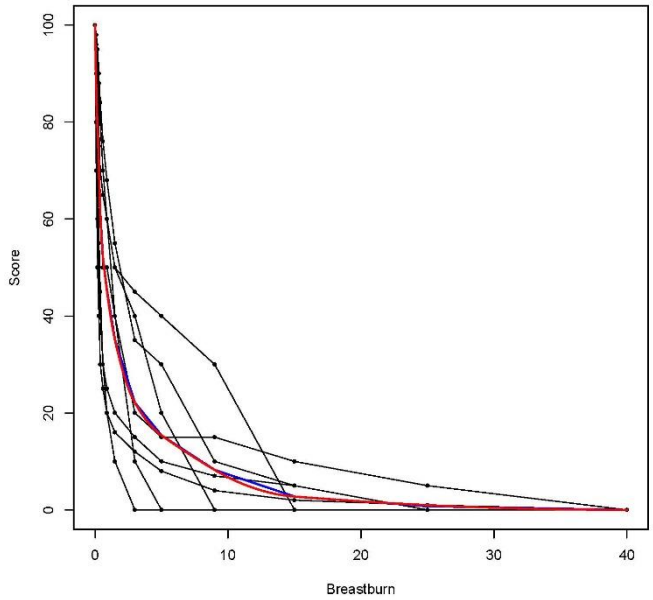


Figure 4.2 *Spline function representing the relationship between the prevalence of breast irritation and the score. The black lines are the scores of the individual experts; the blue line represents the average; the red line represents the spline function derived of the individual scores.*

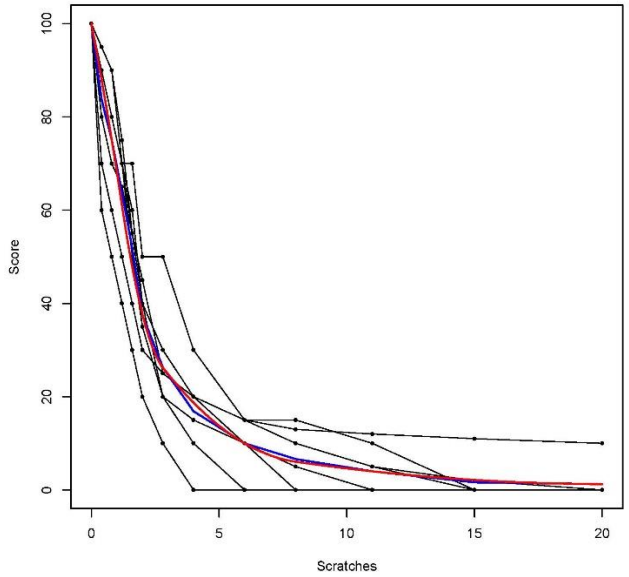


Figure 4.3 *Spline function representing the relationship between the prevalence of scratches and the score. The black lines are the scores of the individual experts; the blue line represents the average; the red line represents the spline function derived of the individual scores.*

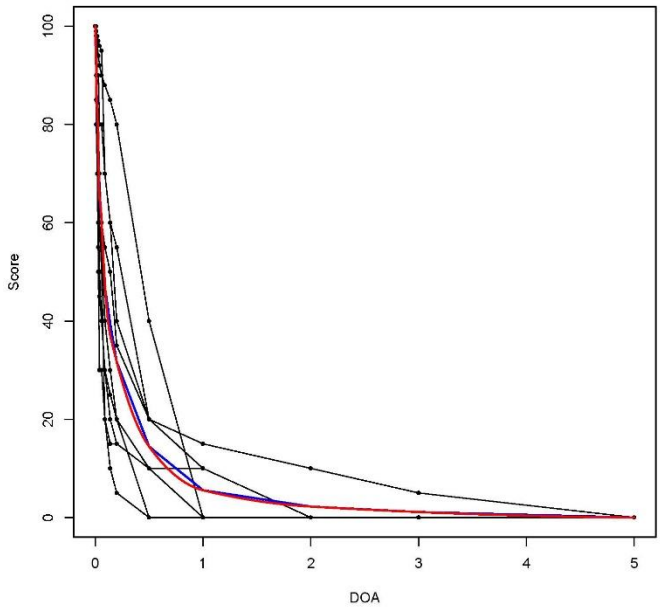


Figure 4.4 *Spline function representing the relationship between the prevalence of Dead-on-Arrival and the score. The black lines are the scores of the individual experts; the blue line represents the average; the red line represents the spline function derived of the individual scores.*

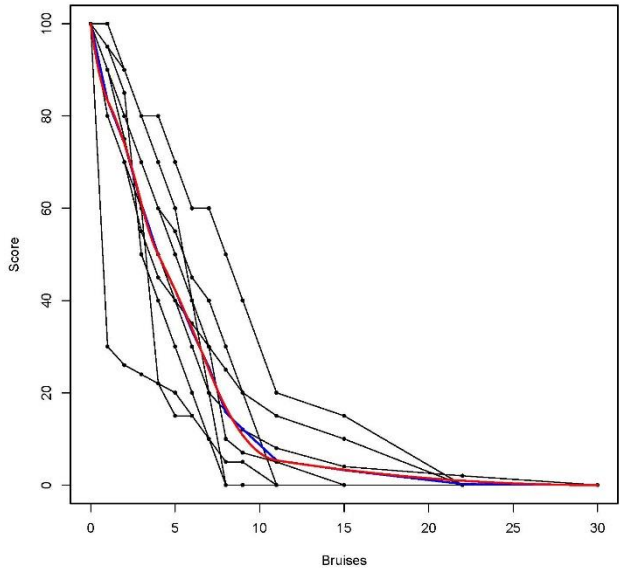


Figure 4.5 *Spline function representing the relationship between the prevalence of Dead-on-Arrival and the score. The black lines are the scores of the individual experts; the blue line represents the average; the red line represents the spline function derived of the individual scores.*

Expert	1	2	3	5	8	9	Score
Early feeding hatchery	70	55	100	65	100	70	**77**
On-farm hatching	100	80	50	100	60	100	**82**
No early feeding	30	20	30	20	0	50	**25**

Figure 4.6 *Decision tree showing the individual expert scores (experts 1-9) and the average score for the absence or presence of early feeding. Note that expert 7 only scored the indicators related to welfare during the end-of-life stage and thus is not included, and that experts 4 and 6 did not assign scores to this indicator.*

Natural light inside the house	Veranda/ outdoor	Number of additional enrichment types in the house	1	2	3	4	5	6	8	9	Score
No	**No**	**0**	20	0	20	40	21	0	0	45	**18**
		1	25	20	30	55	23	20	10	50	**29**
		2	30	40	35	70	25	55	20	55	**41**
		≥ 3	35	50	40	80	27	80	40	60	**51**
	Covered veranda	**0**	35	55	30	70	64	20	30	60	**45**
		1	40	60	40	70	66	55	40	65	**54**
		2	45	75	40	80	68	75	50	70	**63**
		≥ 3	50	75	50	90	70	90	60	75	**70**
	Outdoor range	**0**	65	55	40	70	74	20	50	70	**55**
		1	70	65	50	80	76	55	60	75	**66**
		2	75	75	55	90	78	75	70	80	**75**
		≥ 3	80	75	60	90	80	90	80	85	**80**
	Covered veranda AND outdoor range	**0**	80	60	40	60	84	20	60	80	**60**
		1	85	75	60	70	86	55	70	85	**73**
		2	90	80	70	90	88	75	80	90	**83**
		≥ 3	95	80	80	90	90	90	90	95	**89**

Natural light inside the house	Veranda/ outdoor	Number of additional enrichment types in the house	1	2	3	4	5	6	8	9	Score
Yes	**No**	**0**	25	15	25	70	51	0	10	50	**31**
		1	30	40	30	80	53	20	20	55	**41**
		2	35	55	30	80	55	60	30	60	**51**
		≥ 3	40	55	50	90	57	80	50	65	**61**
	Covered veranda	**0**	40	50	40	80	74	25	40	65	**52**
		1	55	60	45	80	76	60	50	70	**62**
		2	60	80	55	90	78	80	60	75	**72**
		≥ 3	65	80	60	100	80	100	70	80	**79**
	Outdoor range	**0**	70	60	45	80	84	25	60	75	**62**
		1	75	80	70	90	86	60	70	80	**76**
		2	80	85	90	90	88	80	80	85	**85**
		≥ 3	85	85	100	90	90	100	90	90	**91**
	Covered veranda AND outdoor range	**0**	85	75	45	70	94	25	70	85	**69**
		1	90	100	80	90	96	60	80	90	**86**
		2	95	100	100	100	98	80	90	95	**95**
		≥ 3	100	100	100	100	100	100	100	100	**100**

Figure 4.7 *Decision tree showing the individual expert scores (experts 1-9) and the average score for environmental enrichment, natural light and veranda/outdoor range. Note that expert 7 only scored the indicators related to welfare during the end-of-life stage and thus is not included.*

Appendix 5 Histograms of antibiotics usage and mortality

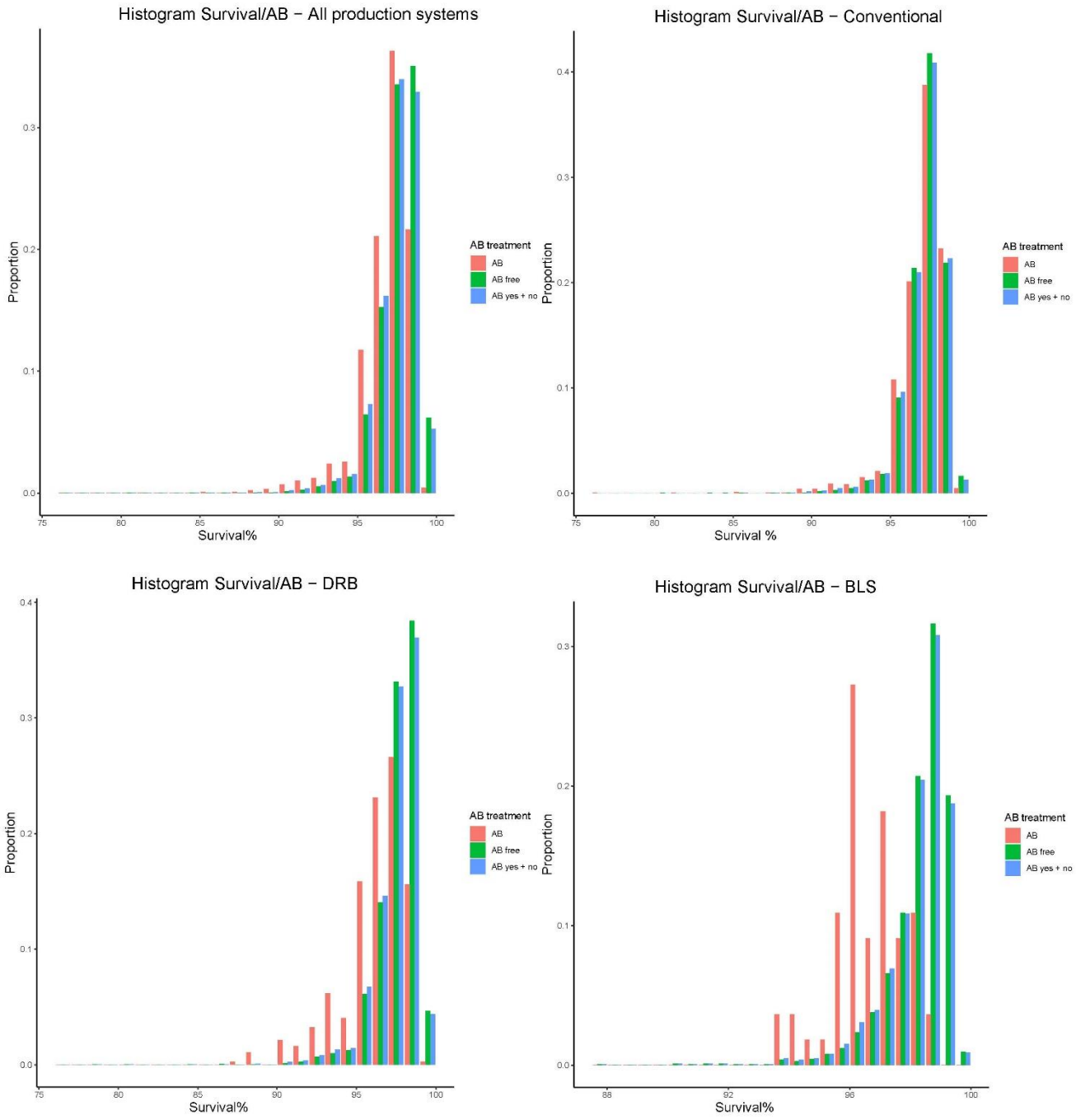


Figure 5.1. *Histograms of survival proportion (X-axis) and flocks separated into three categories per class of survival: no antibiotics used, antibiotics used, or both antibiotics used and not used; histograms are presented for all production systems (upper left), and per production system separately.*

Appendix 6 Locomotion scores

Table 6.1 *Locomotion scores; data from a project partner, that have been collected in experimental pens > 25 m^2. All breeds were tested at a similar stocking density and at similar bodyweight; 4 replicates per breed and 60 birds per pen scored for locomotion.*

A: Locomotion scores per production system (1 trial 2019), shown as % chickens per class for the gait score, and the corresponding Welfare Quality index and score.

	Gait score						Welfare Quality	
Score	0	1	2	3	4	5	Index	Score
Ross 308 d39	1.7	8.3	69.2	20.0	0.8	0.0	95.2	62
Ranger classic d45	5.8	26.7	63.3	4.2	0.0	0.0	99.2	92
Hubbard 987 d45	5.0	20.0	69.2	5.8	0.0	0.0	98.8	89
Ranger Gold d50	5.0	16.7	70.0	8.3	0.0	0.0	98.3	85
Hubbard 957 d50	9.2	33.3	55.8	1.7	0.0	0.0	99.7	97

B:Locomotion scores for 2 conventional breeds (2016-2019), shown as % chickens per class for the gait score, and the corresponding Welfare Quality index and score.

		Gait score						Welfare Quality	
	Score	0	1	2	3	4	5	Index	Score
2016 (n=4 trials)	Ross 308	0.0	1.8	70.3	27.8	0.0	0.0	94.4	58
2018 (n=3 trials)	Ross 308	2.2	9.4	68.6	19.7	0.0	0.0	96.1	68
2019 (n=5 trials)	Ross 308	1.3	9.2	78.2	11.3	0.0	0.0	97.7	80
2016 (n=4 trials)	Another breed	0.4	2.0	65.7	30.6	1.2	0.0	92.7	49
2018 (n=3 trials)	Another breed	1.1	5.3	70.0	22.8	0.8	0.0	94.6	59
2019 (n=5 trials)	Another breed	0.2	0.7	59.4	36.9	2.8	0.0	89.8	39

Appendix 7 Values and scores for the welfare indicators, for 2017 and 2018 separately

Table 7.1 *Values and calculated welfare scores for the broiler (on-farm) stage, based on data collected in 2017 and 2018. Values and calculated scores are presented as the minimum and maximum value or score, and the 0.10, 0.50 and 0.90 percentile to illustrate the distribution of the scores. Table A presents the results for the conventional farms, Table B for the Dutch Retail Broiler and Table C for the Better Life one star production system. Note, that for the values of footpad dermatitis, hock burn, breast irritation and scratches these are presented from 'best' to 'worst', whereas for the scores, this is the other way round, i.e. from 'worst' to 'best' score. Scores and values for stocking density, early feeding and enrichment/natural light/covered veranda and outdoor range are not presented as these are similar as in Tables 6 and 7 in the main text.*

A. Conventional (N= 2693 flocks (2017) and N=2679 flocks (2018))

Variable and score	Year	minimum	P0.10	P0.50	P0.9	maximum
Survival %	2017	80.75	95.70	97.49	98.44	99.85
	2018	76.65	95.55	97.31	98.33	99.85
Mortality score	2017	2.26	44.65	67.38	80.10	98.19
	2018	2.21	42.92	64.98	78.63	98.19
Weighted % footpad lesions	2017	0	40.14	84.28	99.14	100
	2018	0	38.28	81.14	98.86	100
Footpad lesion score	2017	0	13.94	44.39	95.43	100
	2018	0	13.35	39.02	93.93	100
Hock burn %	2017	0	0	7.3	36	88.6
	2018	0	1.3	10	35.6	88
Hock burn score	2017	2.76	9.89	51.77	100	100
	2018	2.81	9.96	40.81	91.97	100
Breast irritation %	2017	0	0	0	0.4	28.8
	2018	0	0	0	0	30
Breast burn score	2017	0.55	61.12	100	100	100
	2018	0.44	100	100	100	100
Scratches %	2017	0	0	0.7	2.5	18
	2018	0	0	0.5	2.25	15
Scratches score	2017	1.42	29.17	78.35	100	100
	2018	2.15	32.87	84.86	100	100
Total Welfare Score	2017	155.40	279.38	382.55	469.17	531.60
	2018	158.06	325.97	445.54	577.39	695.23
Total Score_ABM	2017	112.15	236.13	339.30	425.93	488.35
	2018	114.81	273.09	364.86	435.39	488.95

B. Dutch Retail Broiler (N= 2897 flocks (2017) and N= 2693 flocks (2018))

Variable and score	Year	minimum	P0.10	P0.50	P0.9	maximum
Survival %	2017	80.24	95.96	97.87	98.84	99.92
	2018	76.15	95.67	97.7	98.67	99.99
Mortality score	2017	2.25	47.78	72.47	85.39	99.04
	2018	2.20	44.30	70.19	83.15	99.87
Weighted % footpad lesions	2017	2.14	84.47	99.14	100	100
	2018	0	79.86	99.14	100	100
Footpad lesion score	2017	1.05	44.75	95.43	100	100
	2018	0	37.19	95.43	100	100
Hock burn %	2017	0	1.2	5.2	20.4	85
	2018	0	1	4.8	19	92.5
Hock burn score	2017	3.08	18.80	63.35	92.83	100
	2018	2.43	20.93	65.86	94.56	100
Breast irritation %	2017	0	0	0	0.5	37.7
	2018	0	0	0	0.3	22.2
Breast burn score	2017	0.02	55.53	100	100	100
	2018	1.39	70.37	100	100	100
Scratches %	2017	0	0.5	2	3.8	13.5
	2018	0	0	0.8	2.5	6
Scratches score	2017	2.72	19.86	37.72	84.86	100
	2018	10	29.17	75	100	100
Total Welfare Score	2017	201.60	369.32	446.54	500.26	567.32
	2018	257.71	394.15	471.38	527.82	595.46
Total Score_ABM	2017	113.04	280.38	356.69	410.05	477.01
	2018	169.14	305.39	381.62	437.12	486.94

C. Better Life one star (N=855 flocks (2017) and N=890 flocks (2018))

Variable and score	Year	minimum	P0.10	P0.50	P0.9	maximum
Survival %	2017	87.69	97.022	98.61	99.2	99.93
	2018	90.04	96.624	98.41	99.14	99.7
Mortality score	2017	2.97	61.18	82.36	90.07	99.16
	2018	6.73	56.01	79.70	89.30	96.36
Weighted % footpad lesions	2017	28	87.25	99.43	100	100
	2018	4.71	76.21	98.86	100	100
Footpad lesion score	2017	10.30	50.87	96.95	100	10
	2018	2.24	32.98	93.93	100	100
Hock burn %	2017	0	0	2.33	9	69
	2018	0	0	3	9	38.33
Hock burn score	2017	4.82	44.35	83.52	100	100
	2018	9.47	44.35	78.36	100	100
Breast irritation %	2017	0	0	0	0	24
	2018	0	0	0	0	7.5
Breast burn score	2017	1.13	100	100	100	100

Variable and score	Year	minimum	P0.10	P0.50	P0.9	maximum
	2018	10.76	100	100	100	100
Scratches %	2017	0	0	1.5	4	16.5
	2018	0	0	1	4.5	13
Scratches score	2017	1.71	18.72	50.86	100	100
	2018	2.95	16.09	68.11	100	100
Total Welfare Score	2017	383.43	532.33	607.23	651.83	696.25
	2018	390.51	521.95	603.10	662.47	695.23
Total Score_ABM	2017	177.16	326.06	400.96	445.55	489.97
	2018	184.23	315.67	396.83	456.19	488.95

Table 7.2 *Values and calculated welfare scores for the end-of-life stage, based on data collected in 2017 and 2018. Values and calculated scores are presented as the minimum and maximum value or score, and the 0.10, 0.50 and 0.90 percentile to illustrate the distribution of the scores. Table A presents the results for the conventional farms, Table B for the Dutch Retail Broiler and Table C for the Better Life one star production system. Note, that for the values these are presented from 'best' to 'worst', whereas for the scores, this is the other way round, i.e. from 'worst' to 'best' score.*

A. Conventional (N=2693 flocks (2017) and N=2697 flocks (2018))

Indicator and score	Year	minimum	P0.10	P0.50	P0.9	Maximum
DoA	2017	0	0.02	0.08	0.22	1.26
	2018	0	0.02	0.08	0.23	3.87
DoA score	2017	4.30	30.01	51.40	82.11	100
	2018	0.45	29.21	51.40	82.11	100
Bruises (total %)	2017	0	2.2	5.2	8.67	24.67
	2018	0	2	4.8	8	29.5
Bruises score	2017	0.42	12.55	40.61	71.69	100
	2018	0	16.80	43.81	74	100

B. Dutch Retail Broiler (N=2897 flocks (2017) and N=2693 flocks (2019))

Indicator and score	Year	minimum	P0.10	P0.50	P0.9	Maximum
DoA	2017	0	0.02	0.04	0.11	4.27
	2018	0	0.01	0.03	0.1	2.66
DoA score	2017	0.23	42.09	64.78	82.11	100
	2018	1.45	44.71	73.44	90.22	100
Bruises (total %)	2017	0	3	5	6.84	26
	2018	0	3	4.6	6.5	14.7
Bruises score	2017	0.24	26.44	42.22	61	100
	2018	3.46	29.52	45.37	61	100

C. Better Life one star (N=855 flocks (2017) and N=890 (2018))

Indicator and score	Year	minimum	P0.10	P0.50	P0.9	Maximum
DoA	2017	0	0.01	0.02	0.04	1.62
	2018	0	0	0.02	0.05	2.57
DoA score	2017	3.01	64.78	82.11	90.22	100
	2018	1.54	63.58	82.11	100	100
Bruises (total %)	2017	0	1.5	3	6	12.5
	2018	0	1.5	3	6	14.5
Bruises score	2017	4.52	33.91	61	79.12	100
	2018	3.55	33.91	61	79.12	100

Table 7.3 *Summary of results of statistical analysis comparing the indicator scores between the different systems. Results of the Mood's analysis are based on comparison of the median values of the scores, the Wilcoxon rank-sum test compares the sum of ranks of the systems and tests whether the systems have a similar distribution. The table shows the P-values of the comparison between two systems. Data from 2017. Note: stocking density and enrichment scores were not compared, as these were similar for flocks within a production system.*

	Conventional vs Dutch Retail Broiler		Conventional vs Better Life one star		Dutch Retail Broiler versus Better Life one star	
Indicator	Mood	Wilcoxon	Mood	Wilcoxon	Mood	Wilcoxon
Survival%	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001
Mortality score	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001
Weighted % footpad lesions[1]	P<0.0001	0	P<0.0001	P<0.0001	0.0123	0.0136
Footpad lesion score	P<0.0001	0	P<0.0001	P<0.0001	0.0123	0.01360
Hock burn %	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001
Hock burn score	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001
Breast burn %	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001
Breast burn score	0.0064	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001
Scratches %	P<0.0001	P<0.0001	P<0.0001	P<0.0001	0.2886	P<0.0001
Scratches score	P<0.0001	P<0.0001	P<0.0001	P<0.0001	0.0112	P<0.0001
Early feeding score	P<0.0001	P<0.0001	0	0	0	0
Total Welfare Score	P<0.0001	P<0.0001	P<0.0001	0	P<0.0001	0
Total Score_ABM	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001

[1] First, fractions with score 0, 1 and 2 were calculated. Then, the weighted score was calculated according to the formula:

100 - (2 * fraction score1 * 100 + 7 * fraction score2 * 100)/7 (see [7] page 45).

Table 7.4 *Summary of results of statistical analysis comparing the indicator scores between the different systems. Results of the Mood's analysis are based on comparison of the median values of the scores, the Wilcoxon rank-sum test compares the sum of ranks of the systems and tests whether the systems have a similar distribution. The table shows the P-values of the comparison between two systems. Data from 2018. Note: stocking density and enrichment scores were not compared, as these were similar for flocks within a production system.*

	Conventional vs Dutch Retail Broiler		Conventional vs Better Life one star		Dutch Retail Broiler versus Better Life one star	
Indicator	Mood	Wilcoxon	Mood	Wilcoxon	Mood	Wilcoxon
Survival%	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001
Mortality score	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001
Weighted % footpad lesions[1]	P<0.0001	0	P<0.0001	P<0.0001	0.3379	0.0152
Footpad lesion score	P<0.0001	0	P<0.0001	P<0.0001	0.33791	0.0152
Hock burn %	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001
Hock burn score	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001
Breast burn %	P<0.0001	P<0.0001	0.8375	0.7152	0.0001	0.0004
Breast burn score	0.8485	P<0.0001	P<0.0001	0.7152	P<0.0001	0.0004
Scratches %	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	0.0034
Scratches score	P<0.0001	P<0.0001	P<0.0001	P<0.0001	0.0073	0.0034
Early feeding score	0.0040	0.0034	0	0	0	0
Total Welfare Score	0	0	P<0.0001	0	P<0.0001	0
Total Score_ABM	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001

[1] First, fractions with score 0, 1 and 2 were calculated. Then, the weighted score was calculated according to the formula:

100 - (2 * fraction score1 * 100 + 7 * fraction score2 * 100)/7 (see [7] page 45).

Table 7.5 *Summary of results of statistical analysis comparing the indicator scores between the different systems for Dead-on-Arrival and bruises. Results of the Mood's analysis are based on comparison of the median values of the scores, the Wilcoxon rank-sum test compares the sum of ranks of the systems and tests whether the systems have a similar distribution. The table shows the P-values of the comparison between two systems. Data from 2017.*

	Conventional vs Dutch Retail Broiler		Conventional vs Better Life one star		Dutch Retail Broiler versus Better Life one star	
Indicator	P-value Mood	P-value Wilcoxon	P-value Mood	P-value Wilcoxon	P-value Mood	P-value Wilcoxon
Dead-on-Arrival, %	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001
Dead-on-Arrival, score	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001
Total bruises, %	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001
Total bruises, score	0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001

Table 7.6 *Summary of results of statistical analysis comparing the indicator scores between the different systems for Dead-on-Arrival and bruises. Results of the Mood's analysis are based on comparison of the median values of the scores, the Wilcoxon rank-sum test compares the sum of ranks of the systems and tests whether the systems have a similar distribution. The table shows the P-values of the comparison between two systems. Data from 2018.*

	Conventional vs Dutch Retail Broiler		Conventional vs Better Life one star		Dutch Retail Broiler versus Better Life one star	
Indicator	P-value Mood	P-value Wilcoxon	P-value Mood	P-value Wilcoxon	P-value Mood	P-value Wilcoxon
Dead-on-Arrival, %	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001
Dead-on-Arrival, score	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001	P<0.0001
Total bruises, %	P<0.0001	0.0225	P<0.0001	P<0.0001	P<0.0001	P<0.0001
Total bruises, score	0.0012	0.0236	P<0.0001	P<0.0001	P<0.0001	P<0.0001

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

Wageningen Livestock Research
P.O. Box 338
6700 AH Wageningen
The Netherlands
T +31 (0)317 48 39 53
E info.livestockresearch@wur.nl
www.wur.nl/livestock-research

Wageningen Livestock Research creates science based solutions for a sustainable and profitable livestock sector. Together with our clients, we integrate scientific knowledge and practical experience to develop livestock concepts for future generations.

Wageningen Livestock Research is part of Wageningen University & Research. Together we work on the mission: 'To explore the potential of nature to improve the quality of life'. A staff of 6,500 and 10,000 students from over 100 countries are working worldwide in the domain of healthy food and living environment for governments and the business community-at-large. The strength of Wageningen University & Research lies in its ability to join the forces of specialised research institutes and the university. It also lies in the combined efforts of the various fields of natural and social sciences. This union of expertise leads to scientific breakthroughs that can quickly be put into practice and be incorporated into education. This is the Wageningen Approach.