

# Market driven initiatives can improve broiler welfare – a comparison across five European countries based on the Benchmark method

Peter Sandøe,<sup>\*,†,1</sup> Henning Otte Hansen,<sup>\*</sup> Björn Forkman,<sup>†</sup> Peter van Horne,<sup>‡</sup> Hans Houe,<sup>†</sup>  
Ingrid C. de Jong,<sup>§</sup> Jørgen B. Kjær,<sup>†,#</sup> Søren Saxmose Nielsen,<sup>†</sup> Clare Palmer,<sup>||</sup>  
Helle Lottrup Halkjær Rhode,<sup>†</sup> and Tove Christensen<sup>\*</sup>

<sup>\*</sup>Department of Food and Resource Economics, University of Copenhagen, 1958 Frederiksberg C, Denmark;

<sup>†</sup>Department of Veterinary and Animal Sciences, University of Copenhagen, 1870 Frederiksberg C, Denmark;

<sup>‡</sup>Wageningen University & Research, Wageningen Economic Research, PO Box 8130, EW Wageningen 6700, The Netherlands; <sup>§</sup>Wageningen University & Research, Wageningen Livestock Research, PO Box 338, AH Wageningen 6700, The Netherlands; <sup>#</sup>Friedrich-Loeffler-Institut, Institute of Animal Welfare and Animal Husbandry (ITT), Celle 29223, Germany; and <sup>||</sup>Department of Philosophy, Texas A&M University, College Station, TX 77843, USA

**ABSTRACT** Two kinds of initiatives exist to ensure welfare in broiler production: welfare legislation, where all broiler production in a country or region must comply with legally defined welfare standards; and market driven initiatives, where part of the production must meet specific welfare standards and is sold with a particular label, typically at a price premium, or as part of minimum welfare standards defined by a retailer, a fast-food chain or the like. While the effects of national legislation may be undermined by price competition from lower welfare imported products, the effects of market driven initiatives may be limited by lack of willingness from consumers to pay the extra cost. To investigate how this works out in practice, we compared broiler welfare requirements in 5 European countries, Denmark, Germany, United Kingdom, the Netherlands, and Sweden, in 2018, by means of the Benchmark method. A number of welfare dimensions, covering the input features typically modified in broiler welfare initiatives, were defined. A total of 27 academic welfare experts (response rate

75%) valued the different levels within each dimension on a 0 to 10 scale, and then weighted the relative contribution of each dimension to overall welfare on a 1 to 5 scale. By combining these values and weights with an inventory of existing welfare initiatives, the additional welfare generated by each initiative was calculated. Together with information on national coverage of each initiative, the Benchmark score for each country's production and consumption of chicken meat was calculated. Sweden achieved a much higher Benchmark for national *production* due to higher legal standards than any of the four other countries. The Netherlands, on the other hand, achieved a Benchmark for national *consumption* of chicken at the same level as that found in Sweden, because market driven initiatives complemented more welfare-limited Dutch legislation. So, despite some uncertainties in the Benchmark method, it appears that market driven initiatives can have a strong impact on improving broiler welfare, building on those standards achieved by animal welfare legislation.

**Key words:** broiler welfare, market driven animal welfare, animal welfare legislation, benchmark

2022 Poultry Science 101:101806

<https://doi.org/10.1016/j.psj.2022.101806>

## INTRODUCTION

Two kinds of initiatives exist to counteract animal welfare problems in broiler production. One establishes animal welfare legislation, where all broiler production in the relevant country or region must comply with

legally defined welfare standards. The other consists of market driven initiatives, where part of the production must meet specific welfare standards, typically based on resource-based indicators. These products may subsequently be sold with a particular label, typically at a price premium, or they may be part of minimum welfare standards defined by a retailer, a fast-food chain or the like. However, little is known about the effectiveness, relative and absolute, of such initiatives.

Initiatives to protect the welfare of agricultural animals go back more than 50 yr in Western Europe. Some countries established national welfare legislation, and to avoid unfair price competition the European Community

© 2022 The Authors. Published by Elsevier Inc. on behalf of Poultry Science Association Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Received March 23, 2021.

Accepted February 22, 2022.

<sup>1</sup>Corresponding author: [pes@sund.ku.dk](mailto:pes@sund.ku.dk)

(later the European Union, EU) from the 1990s embarked on the development of minimum requirements for animal welfare standards for some of the animal species most commonly farmed in Europe. However, segments of the European consumer market were still dissatisfied with the minimal requirements defined by animal welfare legislation (Eurobarometer 2007, 2016). Labels on eggs, dairy products and meat that promised welfare provisions going beyond the legally required minima began to emerge in the 1980s. Since then, there has been considerable growth in the market for such products across the Western world (for an overview of this development see Sandøe and Christensen, 2019).

Broilers are the latecomers in this development. Only a few European countries around the turn of this century had national legislation with specific rules regarding broilers, and EU regulation in the form of a directive was agreed in 2007 (European Union, 2007). One of the things regulated is stocking density – the basic allowed stocking density is 33 kg/m<sup>2</sup>, which can be increased up to 42 kg/m<sup>2</sup> under improved management conditions. The directive also, among other things, sets requirements for light intensity, for the duration of periods with darkness, for maximum level of ammonia, and for litter, feeding, and ventilation. These are minimum requirements with which national animal welfare regulation must comply, but individual member countries can set up regulations that go further.

Most EU countries currently regulate broiler production in ways identical with or close to the EU Broiler Directive. However, there are some notable exceptions such as Sweden, where requirements go much further, and where, apparently as a consequence, there is low national broiler production and high imports of broiler meat (Lichter and Kleibrink, 2015).

To enable the improvement of welfare standards without giving those farmers who improve the welfare of their animals a competitive disadvantage, market driven animal welfare has increasingly been viewed as a solution by key stakeholders, including the European Commission (Commission of the European Communities, 2009). This also makes it possible to bring in welfare parameters that have so far not been regulated, such as the genetically determined growth rate (EFSA Panel on Animal Health and Welfare, 2010; de Jong, 2020).

One example of such a market driven initiative can be found in the Netherlands. Here, until around 2012, most broiler production involved fast growing birds and requirements for housing and care that were more or less equivalent to the minimum requirements in the EU directive. Then some of the main retailers changed their supply to sell only fresh meat from broilers that grew slower than the fast growing birds that were reared at a lower stocking density, that were provided with environmental enrichment, and that had a longer uninterrupted dark period (Saatkamp et al., 2019). Within a few years, meat from different varieties of such “welfare” broilers took over the entire market for fresh chicken in the Netherlands, while products from conventional and fast growing birds today still dominate the market for food

service, frozen products, and export (Saatkamp et al., 2019).

What is special about the Dutch situation is that consumers are not given a choice when it comes to fresh broiler meat – only meat based on a production with enhanced welfare is at offer in supermarkets. In other countries, such as Germany, Denmark, and the United Kingdom, welfare labeled meat will often lie next to unlabeled “standard” meat on the cool counter, giving consumers a real choice between welfare chicken with a price premium and standard chicken sold at lower price. Here price competition can still be a big issue.

So, we have seen 2 parallel mechanisms for improving broiler welfare beyond the EU minimal standards: national legislation that goes well beyond it, as in Sweden and to a more limited extent also in other countries, and various market driven animal welfare initiatives that complement the national legislation. To be able to compare the relative effects of these mechanisms, we will apply the so-called Benchmark method (Sandøe et al., 2020).

The key elements of this method are that: 1) it is based on a number of dimensions that influence the welfare of the farm animals in question, in this case broilers; 2) different welfare initiatives – for example a set of legal requirements or requirements for a labeled product – are mapped on to different gradings within these welfare dimensions; 3) expert opinion is used to value and weigh the effect of specific gradings within each dimension of the welfare of the animals; 4) for each welfare initiative Benchmark scores are calculated on a scale from 0 to 100 that indicate the animal welfare provisions of the initiative (where 0 is the lowest and 100 is the highest theoretically conceivable welfare level); and finally 5) by combining information about the Benchmark scores for all initiatives in a country with information about what proportion of the nation’s production or consumption falls under each initiative, the Benchmark score for each country’s production and consumption of broiler meat is calculated on a scale from 0 to 100.

The Benchmark method rests on a number of assumptions, some of which can be challenged. For example, the method takes for granted that the relevant form of animal production complies with the national legislation and the requirements of the various other welfare initiatives in place, which is not always the case (e.g., EUWelfareNet 2014). Also, some of the information used, notably about imports and exports of broiler meat, can be uncertain. However, while we stress these limitations, we still think the method provides a reasonable approximation to an assessment of welfare provisions that, as we will argue in the discussion, can supplement other currently available methods that either face greater limitations or are less feasible.

The Benchmark method has previously been applied to pig production and pig consumption in the same 5 countries as studied here (Sandøe et al., 2020). There it was found that Sweden, which has a level of animal welfare legislation that goes well beyond the minimal EU requirements for pig welfare, could maintain welfare

requirements for the pork consumed that was better than for pork consumed in the other European countries studied. This is despite the fact that Sweden imports a significant amount of pork from countries with lower levels of welfare than in Swedish production. In the present study, we wanted to find out whether that pattern repeats itself in the case of broilers, or whether the relative effects on national welfare provisions of legislation versus market driven initiatives differ in the case of broiler production.

So, the aim of the present study was, by means of the Benchmark method, to measure how the 2 main mechanisms for improving broiler welfare, legislative and market driven, compared in terms of securing welfare provisions of broilers produced and consumed in different countries where these 2 strategies have been pursued to different degrees. Thus, to be able to consider the effect of imports, we not only compared the levels of welfare provisions achieved by different countries in terms of their national *production*, but also the levels achieved for the broiler chickens *consumed* in these countries. The main part of the study focused on the situation in 1 yr, 2018, but we also used the method to document developments in broiler welfare over a 12-yr period in the Netherlands to better understand what has happened there in terms of market-driven animal welfare. Finally, we present data relevant to assess the validity of the applied Benchmark measure, and to document the consistency of expert evaluations.

## MATERIALS AND METHODS

### Selection of Countries

Five Western European countries were selected for this Benchmark study: Denmark, Germany, the Netherlands, Sweden, and the United Kingdom. Details about their production, consumption, import and export of broiler chicken meat, and self-sufficiency rate (production/ consumption) are shown in [Supplementary Table S1](#). Netherlands is the largest net exporter of chicken meat, Germany and Denmark have large exports as well as imports, while broiler production in the UK and Sweden mainly serves their respective home markets, and the 2 countries are net importers of broiler meat.

As members of the EU in the year studied (2018), all countries had to comply with the minimum animal welfare requirement regarding broiler production defined by the EU ([European Union 2007](#)). In addition, one of the countries, Denmark, had legislation that went somewhat further, while a fifth, Sweden, had legislation that went much further.

### Identification of Country Specific Welfare Initiatives

Based on information derived from retailers, producer organizations, public webpages, and other literature in the field, as well as personal contacts with researchers and stakeholders in the respective countries, we created

an overview for each country of initiatives with a reasonable market share that aimed to raise broiler chicken welfare provisions to the level required by the EU. For each country, we included initiatives in the form of national legislative requirements and the main market driven welfare initiatives, including organic production, where we could locate data on required welfare criteria. Smaller initiatives where a few farmers produce chicken with higher welfare were not included in the analysis, because we found that there is a real trade-off between the use of resources to identify and classify all initiatives in a country and the benefits in terms of obtaining a more complete picture of the chicken welfare in the countries. The vast majority of chickens produced according to the experts we consulted are represented in the main 3 to 5 initiatives in a country. The market driven initiatives included both state supported animal welfare labels, and private animal welfare labels, that is, labels managed by companies involved in broiler chicken production, by retailers, by animal welfare NGOs, or through collaboration between some of these. A list of the welfare initiatives for the 5 countries, their market shares, as well as the import and export data used in the analysis can be seen in [Supplementary Table S2](#).

### Definition of Welfare Dimensions

To be able to compare the different initiatives we defined a number of *welfare-dimensions*, for example, stocking density, growth rate and light programmes, most of which were input variables. For each of these we presented different possible gradings. For example, stocking density can range from above 42 kg per m<sup>2</sup>, which does not meet EU requirements, to below 20 kg per m<sup>2</sup> as found in some organic production systems. The dimensions were chosen to cover the most important kinds of welfare requirements adopted by the different broiler chicken welfare initiatives. The starting point for each dimension was the minimum grading (in terms of welfare provisions) of the most common production systems to be found internationally. The other gradings reflected what was typically found in the mentioned welfare initiatives, including the EU minimal requirements, with the highest grading reflecting the current top grade based on the authors' knowledge of broiler chicken production systems in different countries. The aim was to ensure that each broiler chicken welfare initiative could be defined in terms of a combination of grades within the defined welfare-dimensions. In total, 14 dimensions, each with between 2 and 7 gradings, relating to the keeping of broiler chickens, were defined ([Table 1](#)).

### Calculation of Animal Welfare Score Based on Expert Assessments

In order to estimate an animal welfare score for each welfare initiative we used expert assessments. Altogether, 36 recognized European university experts recruited from the network of the researchers involved

**Table 1.** Overview of welfare dimensions and gradings defined in relation to housing and management of broilers.

Dimension	Range of welfare gradings
Stocking density	Four gradings, FROM stocking density above 42 kg per m <sup>2</sup> (does not meet EU requirements) TO stocking density below 20 kg per m <sup>2</sup>
Foot pad dermatitis initiatives	Two gradings, FROM the official veterinarian evaluates and identifies indications of poor welfare conditions such as abnormal gradings of foot pad dermatitis during postmortem inspections (meets EU requirements), leading to severe foot pad dermatitis in about 40% of broiler chickens in 50% of the flocks; TO monitoring postmortem inspections at slaughter, with interventions in cases of high gradings of foot pad dermatitis, leading to severe foot pad dermatitis in about 40% of broiler chickens in 10% of the flocks
Outdoor access	Six gradings, FROM kept indoors with no outdoor access (meets EU requirements); TO access to a veranda and an outside area with vegetation
Maximum flock size in broiler house	Four gradings, FROM no requirement (as per EU standards); TO 1,000 chickens per broiler house
Use of roughage or whole kernels	Three gradings, FROM no roughage provided (compatible with EU standards); TO roughage (e.g., wheat, maize, other roughage) is provided daily
Indoor air quality	Three gradings, FROM no requirements (does not meet EU standards); TO a limit of under 20 ppm NH <sub>3</sub> and/or 3,000 ppm CO <sub>2</sub>
Light programmes (hours of light and darkness)	Four gradings, FROM no requirements regarding the light programme (does not meet EU standards); TO 8 hours of continuous darkness per 24 h
Light intensity	Three gradings, FROM light intensity below 20 lux (does not meet EU standards); TO light intensity above 50 lux covering a minimum of 80% of the area during the photoperiod
The presence of natural sunlight	Two gradings, FROM no sunlight in the broiler house (compatible with EU standards); TO sunlight required in the broiler house
Slaughter ages, taking into account whether the chickens are fast- or slow-growing	Four gradings, FROM use of fast-growing genotypes that are typically slaughtered at 40 d of age (average daily growth rate 60 grams or higher) (as per EU standards); TO use of slow-growing genotypes that are typically slaughtered at an age above 70 d (daily growth rate of 40 grams or less)
Pecking and resting enrichment	Four gradings, FROM no pecking and resting enrichment (compatible with EU standards); TO bales or other pecking enrichment (e.g., pecking stones) required and requirements regarding resting enrichment (perches or platforms)
Thinning (when part of the flock is removed from the broiler house)	Two gradings, FROM thinning is used (meets EU requirements); TO thinning is not used
Catching and loading when all birds are removed from the broiler house	Five gradings, FROM no requirements regarding the catching and loading method (does not meet EU requirements); TO mechanical catching and loading with specifications concerning speed and maximum fall distance and physical requirements
Transportation to slaughter	Seven gradings, FROM the broiler chickens are transported for 24 h without a break, and after 24 h, the broiler chickens have a break of 24 h with food and water, and are transported for a further 24 h (does not meet EU requirements); TO the broiler chickens are not transported to slaughter (mobile slaughtering units are used)

in this study were invited to fill out a questionnaire accessible via a link in an invitation mail. The invited experts came from Austria, Belgium, Denmark, Germany, Finland, France, Italy, the Netherlands, Norway, Spain, Sweden, Switzerland, and the United Kingdom. The first invitation was sent to the experts by email from the first author of the paper on November 12, 2020 and a reminder was sent on November 23, 2020. The questionnaire was closed and the data was harvested on November 30, 2020.

On the first page of the survey the respondents were given the following information: “The aim of the survey is to evaluate the animal welfare effects of various initiatives for broiler production in comparison to the EU standards. The findings based on the responses given by you and a number of other experts will feed into an effort to benchmark broiler chicken welfare across 5 countries. It will take 15 to 20 min to complete the questionnaire. By taking part in this survey you consent that the data from this survey will be used in our studies, and you understand that all analysis and presentation of data from this survey will be anonymous and in no way tracked back to you personally.” After that they were asked: “Do you consent with these conditions?” – and

only if they chose the option “I consent”, were they allowed to continue the survey. The list of the names and email addresses of the invited experts is kept on a password-protected, loggable drive. Only anonymized responses, with no identifying information regarding the experts apart from saying that they are academic experts working the field of animal welfare, are shared with third parties. The ethical standards of the survey study were assessed and approved by the Research Ethics Committee for the Health and Sciences Faculties at the University of Copenhagen, permission number 504-0243/21-5000.

In the questionnaire, the experts were asked, for each dimension, to evaluate the welfare consequence of each grading one by one, on a scale from 0 to 10, while they were told to think of the other welfare dimensions as those of standard production.

The experts were told that the scale should be used as follows:

- 0 represents minimal welfare for the birds in a commercial production system.
- 10 represent the best possible welfare for the birds in a commercial production system.

After that, the experts were asked to assess, on a scale from 1 to 5, for each of the 14 welfare dimensions, what weight it should have when compared to other changes in production that might affect animal welfare. (The text of the questionnaire can be found in [Supplementary Table S3](#)). The choice of scales is of course important, in order to strike a balance between being easy to understand and likely to be used in the same way by different respondents and being nuanced enough to capture differences in values/opinions. We chose the 0 to 10 point scale as it is an often used pain-scale whereas the 1 to 5 scale, ranging from 1-not important to 5-very important, is inspired by the Likert scale often used in sociological and psychological research.

Using the answers from the experts, values for the welfare score of each grade within each welfare dimension, and how much each dimension should contribute to total welfare, were calculated as a weighted sum and rescaled to 0 to 100.

Based on these inputs, we calculated the contribution of each welfare initiative to our broiler chicken welfare Benchmark. This was done on a scale from 0 (= minimal welfare for the birds in a commercial production system) to 100 (= the best possible welfare achievable within commercial broiler chicken production). It should be noted that both the maximum and the minimum are theoretical entities. Depending on how experts use the valuation scale for each welfare dimension the possible scores to be achieved will fall within a narrower range. Only if all experts agree that the best grading in all dimensions deserve a score of 10, and they rate all dimensions to be very important for chicken welfare when compared with other changes in production that might affect welfare, is a score of 100 possible; and only if all the experts agree that the worst in all dimensions deserve a score of 0, is a score of 0 possible. (For further details on these calculations, see [Supplementary Table S4](#)).

## **Benchmarking Production and Consumption**

To be able to calculate a Benchmark score for the broiler chicken production found within a country, we considered how large a share of the production was covered by each initiative in the country. We measured this relative to amount (in metric tons) produced.

Furthermore, we estimated scores for the hypothetical situation that all production in a country is carried out in accordance with either EU minimal requirements or in accordance with national animal welfare legislation, and we assessed the extra Benchmark outcomes in national production created by market driven initiatives.

Besides estimating the Benchmark score for the broiler meat *production* in a country we also estimated the Benchmark score for broiler meat *consumption* in a country. These 2 values can differ considerably. For example, a lower score for consumption than for production may occur if a country has high legislated welfare standards for broiler meat production but at the same time imports large amounts of broiler meat produced

under poorer welfare conditions. Conversely, a higher score for consumption than for production may occur if a country is a net exporter of broiler meat, typically with minimal welfare, but at the same time produces significant amounts of welfare labeled broiler meat for the domestic market.

Finally, for the Netherlands, we calculated how the Benchmark score for production and different segments of consumption developed over the period 2007 to 2019.

Data for production, import and export of broiler meat in a country was extracted from the UN databases, FAO-STAT ([FAO, 2021](#)), COMTRADE ([UN Comtrade, 2021](#)), and from the [EUROSTAT \(2021\)](#) database. When necessary – for example, due to delayed statistics, or due to statistics at too aggregated a level from international databases – data was supplemented with information based on national market observations.

Adjustments to FAOSTAT have been made for import and export values for Denmark and the Netherlands as both countries have a significant amount of imports followed by re-exports. The adjustments were made in order to obtain import and export figures that are linked more directly to the part of production that is exported and to the part of consumption that is imported, rather than being a measure of the amount of chicken meat crossing borders.

Based on consulting Danish experts, we propose that 45% of Danish consumption in 2018 was imported (see [Supplementary Table S2](#)) rather than using figures from FAOSTAT on imported quantity divided by consumed quantity which resulted in an estimated import share of consumption of 91% (see [Supplementary Table S1](#)). And more importantly, rather than using figures from FAOSTAT to estimate Dutch import share of consumption as 250% (see [Supplementary Table S1](#)), a Dutch expert suggested that more likely 28% of Dutch consumption is imported (see [Supplementary Table S2](#)). Furthermore, for Denmark, we propose that instead of 91% of production being exported as can be estimated from the FAOSTAT (see [Supplementary Table S1](#)), we have estimated the export value to be 45% of production (see [Supplementary Table S2](#)) after consulting Danish experts. Similarly, for the Netherlands, instead of 132% of production being exported as can be estimated from the FAOSTAT (see [Supplementary Table S1](#)), we have used a Dutch expert to estimate that more likely 50% of production is exported (see [Supplementary Table S2](#)). The large import and export shares for Denmark and the Netherlands as estimated using FAOSTAT are caused by considerable amounts of chicken meat being re-exported which skewed the picture of the net import and export shares that we focus on here.

Data on market shares of different labels and forms of products in the 5 countries was gathered via contacts with academic experts, civil servants and representatives of slaughterhouses, producer organizations, and retailers in the different countries. Particularly in the case of the market shares of various private welfare labels, accurate information can be difficult to get, but the best possible estimates are presented in [Supplementary Table S2](#).

## Statistical Comparison of Production and Consumption Benchmark Scores

The Benchmark scores for “Production” and “Consumption” between countries were compared using generalized linear models with the `lmer`-function in the `lme4`-package (Bates et al., 2015) in R (R Core Team, 2021). One model included “Production” as the outcome and the other model “Consumption” as the outcome. In both models, “Country” was included as a fixed effect, and “Expert” was included as a random effect, to take into account the score for all countries calculated based on each of the responses of each expert. The difference between countries was compared using the `emmeans`-package (Lenth, 2020), while these pairwise comparisons were adjusted for multiple comparisons using the Tukey posthoc test. An adjusted  $P$ -value of 0.05 was considered statistically significant. The fit of the model was checked by assessing if the standardized residuals were independent, identically distributed Normal  $(0, \sigma^2)$ .

## Assessing the Validity of the Benchmark Method

To be able to assess the validity of the Benchmark method, we compared an existing hierarchy of welfare levels in broiler meat welfare initiatives in Denmark, to see whether these matched the welfare provisions actually measured by the Benchmark approach. One of the Danish initiatives was a governmental animal welfare label, introduced by the Danish Ministry for Food, Agriculture and Fisheries, for broiler meat and broiler meat products in 2018 (Fødevarestyrelsen (No date)). This label has 3 levels, with what is considered to be ascending animal welfare requirements. Another related but slightly different 4-level label, covering broiler meat and broiler meat products, was set up by one of the largest Danish retailers, COOP (No date). By comparing the Benchmark values achieved by the respective levels of these labels, including the organic label, which is widely assumed to have a high level of broiler welfare, the criterion validity of the Benchmark could be estimated, that is, to what extent the sizes the Benchmark values corresponded to the expected relative level of welfare in the labels the Benchmark measured. Also, the Benchmark scores for the governmental label and the retail label were compared with the Benchmark value for Danish standard broiler chicken.

## Assessing the Experts’ Use of the Scale and Their Consistency in Terms of Ranking of Welfare Provisions

To assess the consistency of the experts we carried out 2 kinds of comparisons. First, we compared differences in how the experts used the scale, measured by what, based on their values and weights, would be the minimum and maximum Benchmark that could be achieved,

given their different gradings within the 14 welfare dimensions. Second, we compared the degree to which the experts agreed in their ranking of the 5 countries.

## RESULTS

### Expert Responses

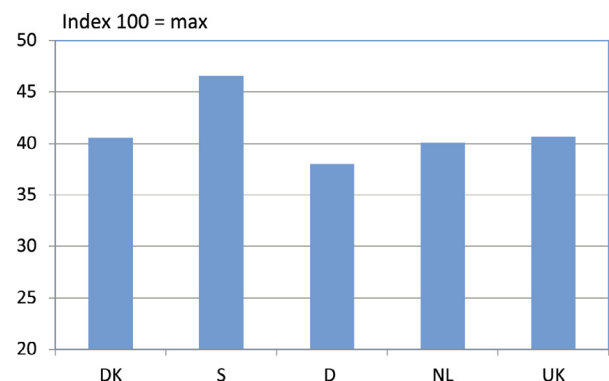
Of the 36 academic experts in broiler welfare whom we asked to contribute to this assessment, 27 had filled out the questionnaire for assigning values and weights when we closed it, after one reminder. This gave a response rate of 75%.

A table with all the expert responses is found as [Supplementary Table S4](#). The variation between experts is further analyzed below.

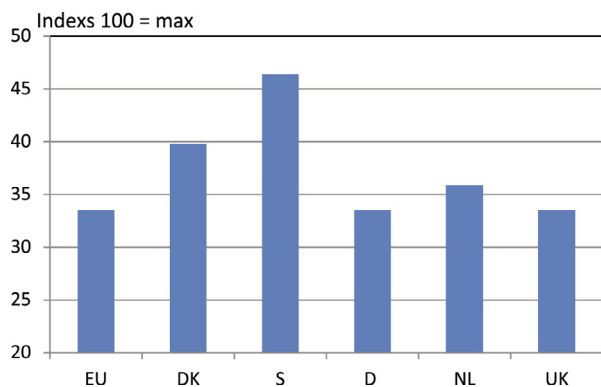
### Benchmark Across Five European Countries

The Benchmark scores for production in the 5 countries (Figure 1), the Benchmark scores for minimum legal requirements in each country (Figure 2), the additional Benchmark scores for market-driven initiatives in each country (Figure 3) and the Benchmark scores for production vs. consumption in each country (Figure 4) are presented below.

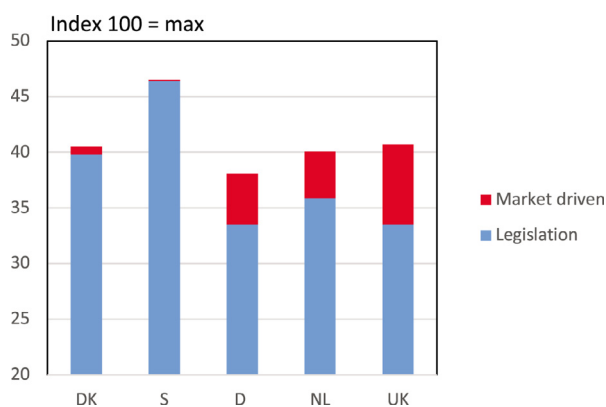
The five countries separated into 3 groups based on the following model: Group I consisted exclusively of Germany, as the Benchmark score for German production (37.5, 95% confidence interval [CI]: 31.1–44.0) was significantly lower than the benchmark scores for all other countries using pairwise comparisons ( $P < 0.0001$ ). The middle group (group II) consisted of the Netherlands, Denmark, and the United Kingdom as a pairwise comparison of Benchmark scores for these 3 countries were not significantly different from each other but the Benchmark scores differed significantly ( $P < 0.0001$ ) from the Benchmark scores for Germany and Sweden, respectively. The production Benchmark scores for group II countries were: the Netherlands (39.6; 95% CI: 33.2–46.1), Denmark (40.0; 95% CI: 33.6–46.4), and the United Kingdom



**Figure 1.** Mean Benchmark scores (on a scale from 0 to 100) for broiler production in Denmark (DK), Germany (D), the Netherlands (NL), Sweden (S), and the United Kingdom (UK). No significant pairwise differences were observed between the Benchmark values of UK, NL, and DK, whereas all other pairwise values were different. The data on which this is based are to be found in Supplementary Tables S1, S2, and S4.



**Figure 2.** Comparison of hypothetical broiler production Benchmark scores (on a scale from 0 to 100) as they would have been if all broiler production followed the EU’s minimal requirements (the EU column) or in accordance with national regulation (including customary standards) (the other columns) in Denmark (DK), Germany (D), the Netherlands (NL), Sweden (S), and the United Kingdom (UK). The data on which this is based are to be found in Supplementary Table S4.

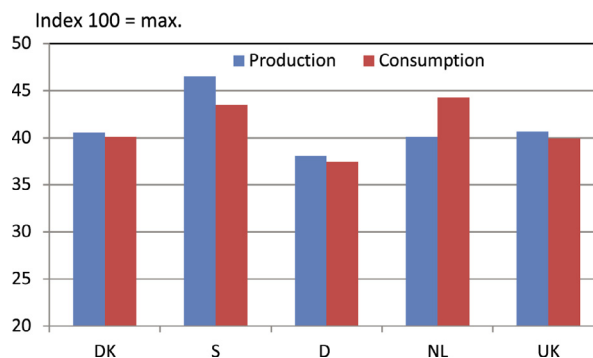


**Figure 3.** The actual broiler production Benchmark scores (on a scale from 0 to 100) including scores from production following legal minimums (blue part) and market driven initiatives (red part) in Denmark (DK), Germany (D), the Netherlands (NL), Sweden (S), and the United Kingdom (UK). The data on which this is based are to be found in Supplementary Tables S1, S2, and S4.

(40.2; 95% CI: 33.7–46.6). Finally, group III consisted of Sweden which had a significantly ( $P < 0.0001$ ) higher Benchmark score for production than the other countries 46.2 (95% CI: 39.8–52.6). The different expert scores are found in [Supplementary Table S5](#).

The differences between the countries may be due to either differences in legislation or differences in market driven initiatives. To see the effect of legislation, [Figure 2](#) illustrates what the national production Benchmark scores in the five countries would have been if all broiler production had followed national regulations (including customary standards). For comparison, we have also added a measure of what the national production Benchmark scores would have been, had all broiler production followed the EU’s minimal requirements.

As can be seen when comparing the Benchmark scores for production as shown in [Figure 1](#) with the Benchmark score for the EU minimum requirements ([European Union, 2007](#)) as shown in [Figure 2](#), all 5 countries achieved a Benchmark score in their broiler production that was between 4.4 and 13.1 Benchmark points higher



**Figure 4.** Comparison of mean Benchmark scores (on a scale from 0 to 100) for broiler production as well as for consumption of chicken meat in Denmark (DK), Germany (D), the Netherlands (NL), Sweden (S), and the United Kingdom (UK). For the consumption Benchmark, no significant pairwise differences were observed between the Benchmark scores of DK and UK and no significant pairwise differences between Benchmark scores of NL and S, whereas all other pairwise comparisons of Benchmark scores for consumption were statistically different. Country-wise, the comparison between Benchmark scores for production and consumption showed significant difference for Germany, Sweden, and the Netherlands but not for Denmark and the United Kingdom. The data on which this is based are to be found in Supplementary Tables S1, S2, and S4.

than the EU minimum requirements. At the same time there were clear differences between the five countries as regards legislation. Two of the countries, Germany and the United Kingdom, had legislation that went no further than EU minimum requirements.

The Netherlands, Sweden, and Denmark went further than the EU demands by requiring that foot pad dermatitis was monitored, and that actions were taken if problems beyond a certain level were found. Denmark also, in line with Sweden, differed from the others by having a customary standard that broiler chickens were not transported for more than 8 h, where the EU allows transport up to 12 h before a mandatory break (after which the transport may continue). Furthermore, the Swedish and Danish chicken farmers had a customary standard regarding acceptable depopulation methods that goes beyond EU requirements. In addition, Sweden had stricter requirements than the other countries in 3 respects: First, a lower stocking density was required, that is, a maximum limit of 36 kg per m<sup>2</sup> as compared to 42 kg/m<sup>2</sup> required by the EU. Second, the Swedish legislation had stricter requirements for indoor air quality in terms of accepted levels of ammonia and carbon dioxide in the barns in comparison to EU requirements. Third, in terms of lighting programmes, Sweden required 6 hours of continuous darkness, while the EU, even though it required a total of 6 h of darkness, only required there to be four hours of *continuous* darkness.

So, the high Swedish Benchmark score for production reflects a higher level of legal requirements compared to that found in the other four countries. Denmark had higher legal requirements than Germany, the Netherlands, and the United Kingdom. However, both the Netherlands and the United Kingdom reach the same Benchmark level as Denmark in their respective broiler production due to larger contributions from market initiatives. This can be seen in [Figure 3](#).

As can be seen from Figure 3, an increase, between 4 and 7 Benchmark points, was generated by market driven animal welfare in the three countries with the lowest legal standards of broiler welfare, Germany, the Netherlands, and the United Kingdom, whereas in the 2 countries with the highest legal standards, Sweden and Denmark, the contribution from market driven animal welfare was minimal.

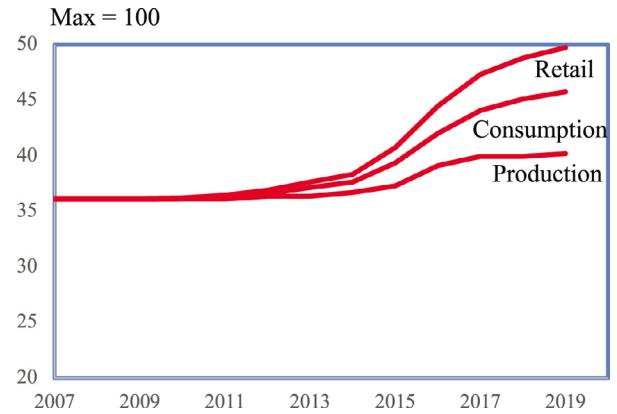
Another important aspect to consider when comparing Benchmark scores between countries concerns consumption within each country. Again, in terms of differences in Benchmark scores for consumption, the countries' Benchmark scores formed 3 groups: i) Germany with a consumption Benchmark score of 36.9 (95% CI: 30.6–43.3) had a Benchmark score that was statistically lower than the consumption Benchmark scores for the other countries ( $P < 0.0001$ ); ii) the Benchmark scores for United Kingdom with 39.4 (95% CI: 33.0–45.8) and Denmark with 39.6 (95% CI: 33.2–46.0) were not statistically different from each other but differed statistically significantly ( $P < 0.0001$ ) from the Benchmarks of the other 3 countries, and finally iii) the Benchmark scores for Sweden with 43.1 (95% CI: 36.7–49.4), and the Netherlands with 44.0 (95% CI: 37.6–50.3) were not statistically different from each other but were statistically higher than the Benchmark scores for the other countries ( $P < 0.0001$ ). For a full list of the expert scores see Supplementary Table S5.

The results of the comparison between the Benchmark scores based on production and consumption for the individual countries are depicted in Figure 4. There was a significant difference between the Benchmark scores for production and consumption for Germany:  $-0.6$  (95% CI:  $-1.0$ ;  $-0.2$ ;  $P = 0.003$ ), Sweden:  $-2.5$  (95% CI:  $-3.1$ ;  $-2.0$ ;  $P < 0.0001$ ), and the Netherlands:  $4.9$  (95% CI:  $4.4$ ;  $5.5$ ;  $P < 0.0001$ ), but not for Denmark ( $0.2$ , 95% CI:  $-0.4$ ;  $0.8$ ;  $P < 0.48$ ) and the United Kingdom ( $-0.2$ , 95% CI:  $-0.7$ ;  $0.4$ ;  $P < 0.56$ ).

These differences indicate that high local legal requirements for broiler production may be counteracted by imports. The Netherlands is, as mentioned, a country with a large and very competitive export of broiler meat. Despite this, the Netherlands also had a local, market-driven initiative for chicken meat with higher required animal welfare provisions for broilers (including a slower growth rate than for standard broilers). As a result, a relatively large Benchmark value, 3.9 Benchmark points higher than that found in Dutch broiler production and 7.9 Benchmark points higher than that required by Dutch legislation, was found for Dutch broiler consumption. The Benchmark level achieved was at the same level as that achieved for broiler consumption in Sweden.

### Mapping of Benchmark Scores Over a Number of Years in One Country

We wanted to better understand what has happened in the Netherlands in terms of market-driven animal



**Figure 5.** Yearly Benchmark scores (on a scale from 0 to 100) in the Netherlands for broiler production, overall consumption of chicken meat and consumption of fresh chicken meat from supermarkets (“Retail”) over the period 2007–2019. Calculations by Peter van Horne from Wageningen Economic Research based on Stichting Avined (2020). Details also provided in S2.

welfare. So, in Figure 5, we have mapped Dutch Benchmark scores over a 12-yr period.

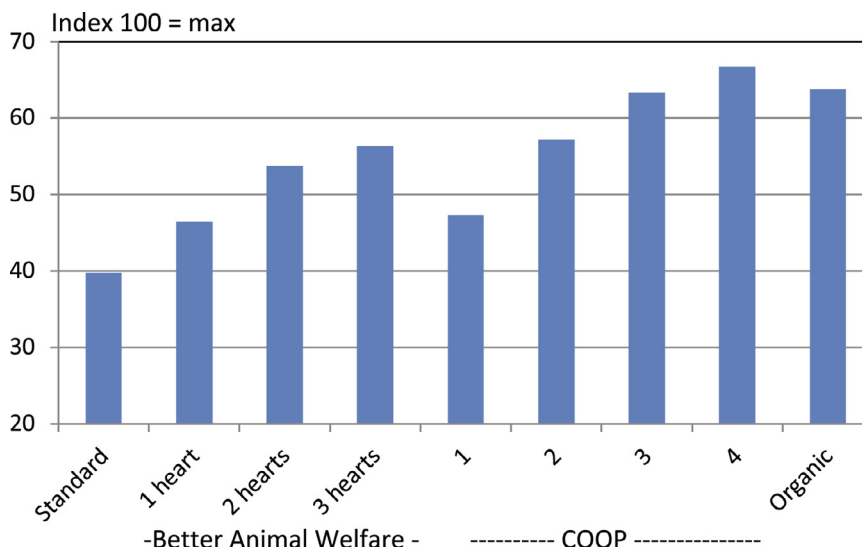
From Figure 5 it can be seen that since 2012, there has been a steady increase in the broiler Benchmark for both production, overall consumption and consumption of fresh chicken meat sold from supermarkets, which seems to plateau around 2018. It also appears that what has driven the development has been a constantly growing national supermarket sale of fresh broiler meat with higher welfare standards. The consumption of frozen and processed chicken products and consumption in the food service sector has been affected very little and was by the end of the period still mainly based on standard broilers; and there was also a significant export market mainly of standard broilers.

### Validity of Broiler Benchmark Scores and Consistency of Expert R improving Levels of Welfare Responses

To assess the criterion validity of the expert opinions, we compared the different forms of Danish broiler production falling under the 2 animal welfare labels as well as organic production applied to chicken meat. These labels are commonly believed to provide increasing levels of welfare compared to standard broiler production undertaken in accordance with Danish legislation, through requiring higher welfare provisions. The results of this comparison are presented in Figure 6.

Both of the main labels, the one defined by the Ministry of Food, Agriculture and Fisheries of Denmark, “Bedre Dyrevelfærd”, used by a range of supermarkets, and the private label defined and used by the supermarket chain COOP, have been designed, based on expert consultations, to have ascending levels of animal welfare. Thus, the first level, which is based on a form of improved indoor production, is supposed to reflect better welfare than the standard production, and for each further level there is supposed to be further improved welfare. The highest level on the state label “Bedre





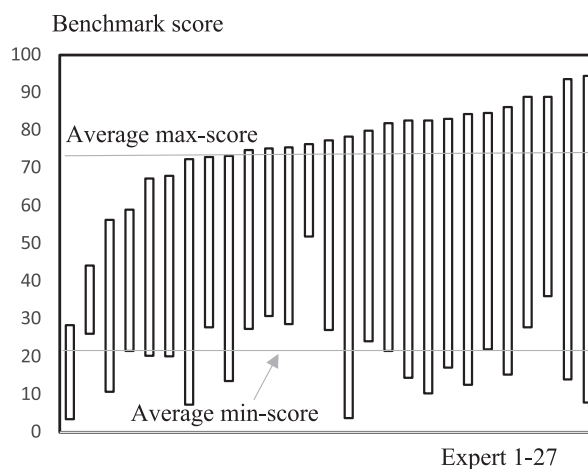
**Figure 6.** The calculated Benchmark-scores (on a scale from 0 to 100) of the different Danish animal welfare initiatives relating to broilers: “Standard” covers unlabelled broiler products produced according to Danish legislation. One to 3 hearts are three levels of the official welfare label “Bedre Dyrevelfærd” (“Better Animal Welfare”), Coop 1-4 are four levels of the retail welfare label “Dyrevelfærdshjertet”, “Organic” represents organic broiler products. The data on which this is based are to be found in Supplementary Tables S2 and S4.

Dyrevelfærd” requires outdoor access. The first level in the COOP label requires improved indoor conditions while the remaining three levels of the COOP label require outdoor access. The highest level in the COOP label requires organic production and some extra initiatives beyond that. (The criteria for the labels can be seen in [Supplementary Table S4](#).) Comparing the Benchmark scores with what could be expected, there are only some very small deviations. Such minor differences may be due to differences in how the requirements are specified. However, overall there seems to be a very good correspondence between the Benchmark scores and what would be expected from the way the labels have been defined.

Another way to approach the validity of the Benchmark score is by looking at the diversity of responses among the experts involved. First, in [Figure 7](#), we look at the way the experts used the scale in that, for each expert, we calculate their range of possible Benchmark scores given the gradings presented within the different welfare dimensions.

As can be seen from [Figure 7](#) there are major differences between the experts, both in terms of how large a part of the scale they were using when valuing the various gradings within the welfare dimensions and in terms of where on the scale their answers were placed. This has consequences for the possible range of scores an initiative can achieve. Thus, with the current gradings and the current experts the highest Benchmark value achievable is 72.9 and lowest possible value is 21.4. The difference between 72.9 and the theoretical maximum of 100 represents the difference between what it is currently possible to achieve in terms of broiler welfare according to the experts and what, according to the experts, can be considered as ideal broiler welfare.

Even though the experts are using the scales differently, they can still be consistent in the way that, based on their evaluations, they rank different animal welfare



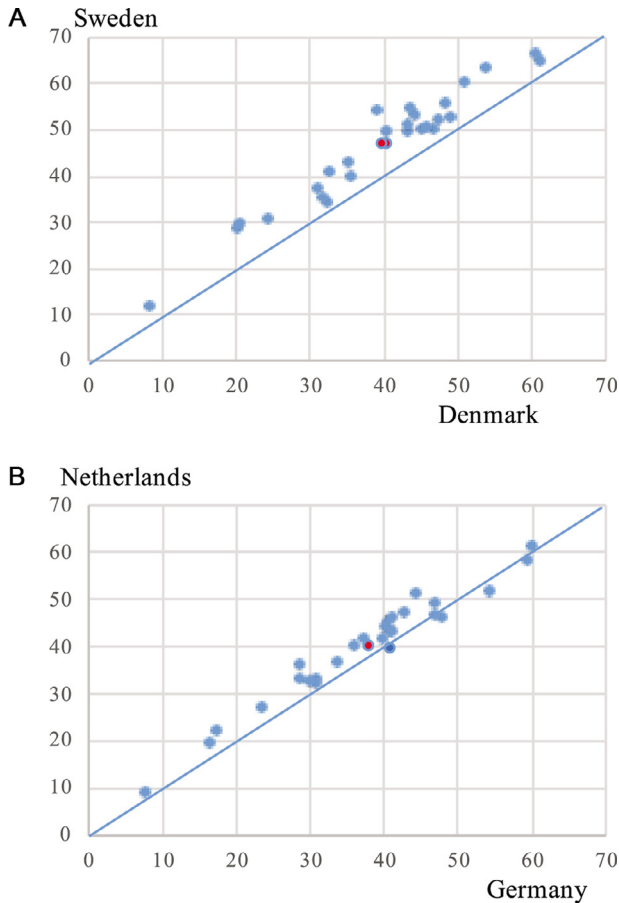
**Figure 7.** The range of scores across welfare dimensions based on responses from the 27 experts. Also the average maximum scores and the average minimum scores are calculated and depicted as horizontal lines. The data on which this is based are to be found in Supplementary Table S4.

initiatives. As can be seen from [Supplementary Table S5](#), Sweden is ranked best among all 27 experts. Also, it can be seen that the differences among D, DK, NL, and UK are minimal. These results have been illustrated in [Figures 8A](#) and [8B](#).

As can be seen, although the experts varied hugely in their range of Benchmark scores, the resulting Benchmark scores were still very consistent when comparing countries, illustrated by the almost perfect correlation between the Benchmark scores in the pairwise comparison of countries.

## DISCUSSION

The results presented in this paper concern 5 European countries that, in 2018, were all subject to the



**Figure 8.** (A) Comparing Benchmark scores for broiler production by all 27 experts regarding Sweden and Denmark. The red dot represents the average score among experts. The data on which this is based is to be found in Supplementary Table S4. (B) Comparing Benchmark scores for broiler production by all 27 experts regarding the Netherlands and Germany. The red dot represents the average score among experts. The data on which this is based is to be found in Supplementary Table S4.

same minimum requirements for animal welfare as defined by the EU. At the same time, there were major differences between the levels of broiler welfare requirements in these countries, and data presented here showed that 2 main forces were at play: National legislation going beyond the minimum EU requirements and voluntary, market driven initiatives.

We show that animal welfare legislation has a significant effect on the Benchmark measure of welfare provisions for broiler chicken *produced* in all 5 countries and that the country with the most far reaching legislation for broiler welfare, Sweden, also reaches a much higher Benchmark score than the other countries. However, we also show that market driven initiatives can have a large effect; this can be seen by comparing Denmark to the Netherlands and the United Kingdom. Denmark had higher legally defined welfare standards compared to the other 2 countries, but ended up with the same Benchmark level of welfare provisions in its national production as The Netherlands and the United Kingdom because market driven initiatives have larger impacts in these countries. Furthermore, market driven initiatives

can have an even larger effect on the Benchmark level of welfare provisions for broilers *consumed* in a country.

It is worth discussing whether it is relevant to look at national *consumption* as well as production when comparing required welfare provisions in different countries (which brings the Netherlands to the same level as Sweden) or whether the focus should only be on national *production* (which will give Sweden an advantage over the Netherlands). We see some value in the first stance. In doing so, we have been inspired by the seminal 2009 report from the British Farm Animal Welfare Council, “Farm Animal Welfare in Great Britain: Past, Present, and Future” (Farm Animal, 2009). This report describes a development where efforts to improve animal welfare by means of unilateral national initiatives led to problems with imports from countries with lower standards, and it is claimed that “the power of the concerned consumer (through market demand) is mostly unrealized”. Since then much has happened in terms of market driven initiatives. However, it has so far been very difficult to assess whether these initiatives made a real difference in terms of animal welfare provisions, or whether they mainly served as window dressing for a system dominated by standard products that reach only a minimal level of animal welfare. In our study, we have shown that market driven initiatives most likely can make a considerable difference.

Of course, it may be argued that the Dutch position – with high welfare requirements in production aimed at the domestic market, and a high export of broilers produced according to the legal requirements for animal welfare – is a bit hypocritical. However, in our view, this should also be seen across a longer time perspective. Market-driven initiatives like those found in the Netherlands may spread to other countries, which already seem to be happening; and this may lead to growing exports of Dutch welfare chicken.

There are at least 2 other initiatives that make large scale comparisons of animal welfare provisions, but neither deliver results with the same focus and specificity as those we deliver here. One initiative is *The Business Benchmark on Farm Animal Welfare* (Amos et al., 2018) and another is *The Animal Protection Index* (World Animal Protection, 2020). These have been presented and discussed in detail in Sandøe et al. (2020). Compared to *The Business Benchmark on Farm Animal Welfare*, which is focused on benchmarking food companies, our Benchmark enables a comparison of welfare scores across countries rather than companies, and it measures the relative contribution of what is done to improve animal welfare. Compared to *The Animal Protection Index*, which is focused on legislation and state policy, our proposed Benchmark is at the same time broader, in that it not only includes animal welfare legislation and other state driven policies but also private initiatives, and more focused in that it allows a comprehensive assessment of a specific field of animal production and consumption, in this case of broilers.

Still, it is reasonable to ask whether the method chosen here to assess welfare, based on expert valuations and weightings, is the best possible.

For instance, a single benchmark score that is based on the sum of scores of individual parameters builds on an assumption that bad welfare in one dimension can be compensated for by good welfare in another dimension. This could be regarded as problematic. On the other hand, having a weight for each dimension allowed us to acknowledge that some dimensions might have a greater impact on animal welfare than others in this compensation process.

Assumptions such as this one about welfare compensation are of course debatable (and they may even be controversial) but, we should note, such assumptions and comparisons are made all the time – though, in our view, not as transparently as here.

There are 2 weaknesses of the Benchmark method compared to the protocols for assessing animal welfare emerging from the EU funded project Welfare Quality (WQ) which also includes a protocol for assessing broiler welfare (Welfare Quality, 2009). The Benchmark primarily looks at resource-based measures of animal welfare – “welfare provisions”, what is provided to the animals in terms of space, enrichment etc. – while WQ looks at the actual reactions of the animals to the housing and management they are subjected to, measured in terms of so-called outcome- or animal-based measures. It might reasonably be argued that direct welfare measures of animals’ reactions give a more accurate account of welfare than more indirect resource-based measures. However, the difference between these methods is less than it appears at first sight. In the Benchmark approach, measures are scored by experts who can translate the resource-based measures to proxy outcomes for the birds. This means that there will also be a focus on outcomes in the Benchmark. However, it is still true that the experts only report about *typical* welfare outcomes, not about variations in outcomes across farms or countries due to differences in the quality of management (de Jong and van Riel, 2020), a feature of the WQ approach. A second disadvantage of the Benchmark approach is that it uses data from welfare legislation and market initiatives, and presupposes that farms are complying with these, rather than investigating how far there is actually compliance on the ground. In practice, there may be only partial compliance with welfare legislation and market initiatives, and the degree of compliance may also vary between countries (EUWelNet 2014).

Despite this theoretical advantage of the WQ approach, in practice this ambitious approach has not so far provided large-scale comparisons. Based on a review of the literature there seems to be at least 3 reasons why WQ so far has not succeeded in implementing an alternative system for measuring welfare outcomes for broilers in different countries and in different housing systems:

First, using the WQ system is time-consuming and requires expertise on the ground De Jong et al. (2016).

have shown that it may be possible to simplify the protocol and thereby make its application less costly. This is surely a promising strategy, but the authors conclude that the proposed “simplification strategies should, however, be validated further, and tested on farms with a wide distribution across the different welfare categories of WQ”. And this work has so far not been undertaken.

Second, the aggregation system found in the current WQ system has, as has been argued by Buijs et al. (2017), some severe limitations when it comes to assessing broiler welfare. Thus, as they argue, “95% of the flocks’ overall classification was explained by 2 measures only (‘drinker space’ and ‘stocking density’)”, both of which are resource based measures, even though, as mentioned above, the basic idea of WQ is to rely primarily on outcome based measures of animal welfare. So, until these problems, and other more general problems relating to the aggregation system of WQ (Sandøe et al., 2019) are solved, it seems that the WQ system has some limitations when it comes to comparing welfare outcomes of different welfare initiatives.

In the Benchmark system, on the other hand, the complex effects of the WQ aggregation system, including those described by Buijs et al. (2017) seem to be avoided. Thus, all inputs from all measures and all selected criteria contribute to the outcome in an additive way to create an index according to the weights given by experts, and no specific measures dominate in the accumulated index.

The third and last reason is, as argued by Sandøe et al. (2020), that there are currently legal limitations to random sampling of farms to be assessed by means of the WQ systems. Without such a sampling, the WQ protocol cannot deliver a representative picture of the welfare of broilers in a specific country.

Another interesting effort to compare broiler welfare across four very different countries, and based on outcome-based measures, is found in Tallentire et al. (2018). Based on data on stocking density, mortality, and carcass condemnation rate from conventional chicken meat production systems, the authors made an assessment of overall animal welfare impact per kg of chicken meat produced in four European countries. This approach, however, had some obvious limitations. First, due to reasons of confidentiality, the names of the countries could not be revealed. Second, only “conventional”, intensive broiler production was assessed and therefore no comparison with market driven alternatives was possible. Third, only a few, crude, outcome-based measures such as mortality and condemnation rate were used. Finally, the main predictor for welfare level seemed to be stocking density (which is also included in the Benchmark system) whereas in the Benchmark the score is an additive function of multiple dimensions.

The current application of the Benchmark method to broiler production also has several advantages over a previous application of the Benchmark to pig production (Sandøe et al., 2020). First, more experts were used in the current study than in previous study. Second, the method was also used to compare the welfare provisions

for broilers over time in one country. Third, results about the maximum and minimum scores achievable have been included.

Regarding the latter point, a possible concern is that we could have re-scaled our results, so that 0 equals the minimum achievable on the current scale (21.4) and 100 equals the maximum achievable on the current scale (72.9). However, we have chosen not to do so because we think the current scale gives a fairer picture of the welfare level as viewed by the experts. When the experts do not use the lowest end of the scale, it seems to reflect their view that the worst option offered is not the worst option possible; and when the experts do not use the highest end of the scale it seems to reflect their view that even with the best options offered, the resulting welfare is not ideal. One caveat here is, of course, that when it comes to the use of the scale the experts vary a lot – from using 18% to using 87% of the scale.

The large variation in the use of the scale does not reappear when it comes to the expert ranking of Benchmark scores for production at national levels. Here there is high consistency which means that, to a significant extent, the experts agree when it comes to the kind of comparisons for which the Benchmark is being used in the current study.

Of course, the approach still has its limitations which have been described above. The main limitation is that experts only report *typical* welfare outcomes, not taking management or compliance with standards into account. Also, there is some uncertainty about the effects of imports and exports of broiler meat. Even though we have tried, on a case-by-case basis, with some help from local experts, to assess the welfare provisions by which imported or exported broiler meat is produced, uncertainty remains. For example, some findings indicate that broiler meat imported from some countries outside the EU may be produced with reasonably high welfare standards (Tuytens et al., 2015; Vissers et al., 2019).

The most important finding of our study is that not only animal welfare legislation but also market driven initiatives can have an important impact on broiler welfare provisions. While stringent national legislation can secure the highest level of welfare provisions in national broiler production, market driven initiatives can also have a significant positive effect. When it comes to national consumption, market driven initiatives can have an effect comparable to that of stringent national legislation.

## ACKNOWLEDGMENTS

The initial phases of the research presented here were funded by the Danish Knowledge Center for Animal Welfare, Danish Veterinary and Food Administration. The authors are extremely grateful for information on national labels, legislation and market shares to Margit Beck, MEG – Marktinfo Eier & Geflügel; Lotta Berg, Swedish Agricultural University; Jason Gittins, Agriculture and Land Management, ADAS; Anina Kjær, Danish Agriculture and Food Council; Victoria Sandilands, Scotland's

Rural College; and a number of other persons from academia, civil service, agricultural organizations, retail, and animal welfare NGOs who have provided data on animal welfare initiatives, market shares, and the like in the countries covered. We are also very grateful to the 27 anonymous academic experts who filled out the questionnaire based on which we calculated the welfare impact of each welfare initiative. And we are grateful to two anonymous reviewers providing useful comments to an earlier version of this paper. Finally, we would like to thank Eliza Ruiz Izaguirre for helping us editing the manuscript.

The research presented here was partly funded by the Danish Knowledge Center for Animal Welfare, Danish Veterinary and Food Administration. The funder had no role in the design of the study, in the collection, analyses, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

## DISCLOSURES

The authors declare that they have no known competing financial interests or personal relationships that influenced or could have appeared to influence the work reported in this paper.

## SUPPLEMENTARY MATERIALS

Supplementary material associated with this article can be found in the online version at [doi:10.1016/j.psj.2022.101806](https://doi.org/10.1016/j.psj.2022.101806).

## REFERENCES

- Amos, N. R. Sullivan, H. van de Weerd. 2018. The business benchmark on farm animal welfare methodology report. Accessed Mar. 2021. [https://www.bbfaw.com/media/1555/bbfaw\\_methodology\\_report\\_2018.pdf](https://www.bbfaw.com/media/1555/bbfaw_methodology_report_2018.pdf)
- Bates, D., M. Maechler, B. Bolker, and S. Walker. 2015. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67:1–48.
- Buijs, S., B. Ampe, and F. A. M. Tuytens. 2017. Sensitivity of the Welfare Quality® broiler chicken protocol to differences between intensively reared indoor flocks: which factors explain overall classification? *Animal* 11:244–253.
- Commission of the European Communities. 2009. Options for animal welfare labelling and the establishment of a European Network of Reference Centres for the protection and welfare of animals. Report from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions COM(2009) 584 final.
- COOP. No date. Dyrevelværdshjertet. Accessed Mar. 2021. <http://dyrevelfaerd.coop.dk/dyrevelfaerdshjertet/>
- de Jong, I. C. 2020. Welfare issues in poultry housing and management: broilers. *Understanding the Behaviour and Improving the Welfare of Chickens*. C. Nicol, ed. Burleigh Dodds Science Publishers, Cambridge, UK.
- de Jong, I. C., V. A. Hindle, A. Butterworth, B. Engel, P. Ferrari, H. Gummink, T. Perez Moya, F. A. M. Tuytens, and C. G. van Reenen. 2016. Simplifying the Welfare Quality® assessment protocol for broiler chicken welfare. *Animal* 10(Suppl. 1):117–127.
- de Jong, I. C., and J. W. van Riel. 2020. Relative contribution of production chain phases to health and performance of broiler chickens: a field study. *Poult. Sci.* 99:179–188.
- EFSA Panel on Animal Health and Welfare. 2010. Scientific opinion on the influence of genetic parameters on the welfare and the resistance to stress of commercial broilers. *EFSA J.* 8:1666.

- Eurobarometer. 2007. Attitudes of EU citizens towards animal welfare. Special Eurobarometer 270. Accessed Sept. 2021. <https://europa.eu/eurobarometer/surveys/detail/470>
- Eurobarometer. 2016. Attitudes of EU citizens towards animal welfare. Special Eurobarometer 442. Accessed Sept. 2021. <https://op.europa.eu/en/publication-detail/-/publication/9bc3a0b7-ec17-11e5-8a81-01aa75ed71a1>
- European Union. 2007. COUNCIL DIRECTIVE 2007/43/EC of 28 June 2007 laying down minimum rules for the protection of chickens kept for meat production. Accessed Mar. 2021. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32007L0043&from=EN>
- Eurostat. 2021. Database. Accessed Mar. 2021. <https://ec.europa.eu/eurostat/data/database>
- EUWelNet. 2014. Report on the main problem areas and their sensitivity to be addressed by knowledge transfer for each of the specific aspects of the legislation chosen for this project. Projec Report. Accessed Jan. 22. <https://edepot.wur.nl/293527>
- FAO. 2021. FAOSTAT. Accessed Mar. 2021. <http://www.fao.org/faostat/en/#data>
- Farm Animal Welfare Council (2009). Farm Animal Welfare in Great Britain: Past, Present and Future. Accessed March 2022. [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/319292/Farm\\_Animal\\_Welfare\\_in\\_Great\\_Britain\\_-\\_Past\\_\\_Present\\_and\\_Future.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/319292/Farm_Animal_Welfare_in_Great_Britain_-_Past__Present_and_Future.pdf)
- Fødevarestyrelsen. No date. The governmental animal-welfare label. Accessed Mar. 2021. <https://www.foedevarestyrelsen.dk/kam-pagner/Bedre-dyrevelfaerd/Sider/Information-in-English.aspx>
- Lenth, R. V. 2020. emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.4.6. Accessed Mar. 2021. <https://CRAN.R-project.org/package=emmeans>
- Lichter, J., and J. Kleibrink. 2015. Geflügelwirtschaft Weltweit - Deutschland im internationalen Vergleich. Handelsblatt Research Institute, Duesseldorf, Germany.
- R Core Team. 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Accessed Mar. 2021. <https://www.R-project.org/>.
- Saatkamp, H. W., L. S. M. Vissers, P. L. M. van Horne, and I. de Jong. 2019. Transition from conventional broiler meat to meat from production concepts with higher animal welfare: experiences from The Netherlands. *Animals* 9:483.
- Sandøe, P.; T. Christensen. 2019. Farm animal welfare in Europe: from legislation to labelling. Working paper. Accessed Mar. 2021. [https://dyreetik.ku.dk/dokumenter/forskningsprojekter/From\\_legislation\\_to\\_labelling.pdf](https://dyreetik.ku.dk/dokumenter/forskningsprojekter/From_legislation_to_labelling.pdf)
- Sandøe, P., S. A. Corr, T. B. Lund, and B. Forkman. 2019. Aggregating animal welfare indicators – can it be done in a transparent and ethically robust way. *Anim. Welf.* 28:67–76.
- Sandøe, P., H. O. Hansen, H. L. H. Rhode, H. Houe, C. Palmer, B. Forkman, and T. Christensen. 2020. Benchmarking farm animal welfare — A novel tool for cross-country comparison applied to pig production and pork consumption. *Animals* 10:955.
- Stichting Avined. 2020. Antibioticumgebruik pluimvee sector in 2019. Avined report, Nieuwegein. (in Dutch). Accessed Mar. 2021. <https://www.avined.nl/themas/pluimveegezondheid/antibiotica>
- Tallentire, C. W., S. A. Edwards, and T. Van Limbergen. 2018. The challenge of incorporating animal welfare in a social life cycle assessment model of European chicken production. *Int. J. Life Cycle Assess.* 24:1093–1104.
- Tuytens, F. A. M., J. F. Federici, R. F. Vanderhasselt, K. Goethals, L. Duchateau, E. C. O. Sans, and C. F. M. Molento. 2015. Assessment of welfare of Brazilian and Belgian broiler flocks using the Welfare Quality protocol. *Poult. Sci.* 94:1758–1766.
- UN Comtrade. 2021. UN comtrade database. Accessed Mar. 2021. <https://comtrade.un.org/data/>
- Vissers, L. S. M., I. C. de Jong, P. L. M. van Horne, and H. W. Saatkamp. 2019. Global prospects of the cost-efficiency of broiler welfare in middle-segment production systems. *Animals* 9:473.
- Welfare Quality. 2009. The Welfare Quality assessment protocol for poultry (broilers, laying hens). The Welfare Quality Consortium.
- World Animal Protection. 2020. Animal protection index. Accessed on Mar. 2021. <https://api.worldanimalprotection.org/#>