Associations between qualitative behaviour assessments and measures of leg health, fear and mortality in Norwegian broiler chicken flocks

Karianne Muri[1]*, Solveig Marie Stubsjøen[2], Guro Vasdal[3], Randi Oppermann Moe[1] and Erik Georg Granquist[1]

[1]Norwegian University of Life Sciences, Faculty of Veterinary Medicine,
PO Box 8146 dep., 0033 Oslo, Norway

[2]Norwegian Veterinary Institute, Department of Animal Health and Food Safety, Section for Animal Health, Wildlife and Welfare, P.O. Box 750, Sentrum, N-0106 Oslo,
Norway

[3]Norwegian Meat and Poultry Research Centre, Lorenveien 38, 0513 Oslo, Norway

*Corresponding author; Karianne Muri, karianne.muri@nmbu.no

Abstract

Qualitative behavioural assessments (QBA) is an animal-based welfare measure that has been included in several on-farm welfare assessment protocols, including the Welfare Quality® (WQ) protocol for poultry. However, there is a scarcity of information about how it relates to other animal-based welfare indicators. The aim of this study was therefore to investigate the associations between QBA and selected animal-based welfare indicators commonly used for the assessment of broiler chicken welfare, i.e. lameness, foot pad dermatitis (FPD), fear of humans (touch test), and mortality.

A total of 50 commercial broiler chicken farms were visited by one observer who conducted on-farm welfare assessments using the WQ protocol. Assessments were done close to the time of slaughter (between day 27 and 34). QBA was analysed using principal component analysis (PCA), revealing two main components, labelled *arousal* (PC1) and *mood* (PC2). The scores for the other welfare indicators were categorised into dichotomous (touch test) or ordinal scales (gait score, footpad dermatitis score and mortality) to deal with skewed distributions caused by homogenous data. To investigate the associations between QBA and the other welfare indicators, we ran logistic and ordinal logistic regression models with these welfare measures as outcomes, and the two components of QBA as the predictors.

Significant negative associations were found between both components of QBA and the chickens' fear of humans, as measured using the touch test. In other words, flocks with higher scores on both *mood* and *arousal* were less likely to have any chickens that were possible to touch by the assessor. A possible interpretation of these associations is that both QBA components may indicate greater liveliness in birds that did not accept to be touched by the observer. Flocks with a higher arousal score, as measured by the first component of QBA (PC1), were also less likely to be in a higher mortality category. For the other selected animal-based measures, there were no associations with QBA. We conclude that QBA needs thorough validation for the routine use in the assessment of broiler chicken welfare, but that the method may provide useful supplementary information in overall welfare assessments. This information may be particularly valuable in a production system, like the broiler industry, where management is highly standardised, sometimes resulting in little between-flock variation in other welfare measures.

Keywords: broiler chickens; fear; footpad dermatitis; lameness, mortality; qualitative behavioural assessments (QBA)

1. Introduction

There is a strong public concern for the welfare of chickens kept for meat production (European Food Safety Authority, 2012; Scientific Committee on Animal Health and Animal Welfare, 2000) and therefore a need for valid, reliable and feasible methods for assessing broiler welfare on farms. On-farm assessments supplements data routinely registered during the production phase and at the time of slaughter, like mortality rate and leg health indicators. The Welfare Quality® (WQ) project developed an on-farm welfare assessment protocol for poultry, which provides detailed systems for assessing the welfare of laying hens and broiler chickens (WelfareQuality®, 2009). Qualitative behavioural assessment (QBA) is one animal-based welfare indicator that has been included in several on-farm welfare assessment protocols, including the WQ protocols. QBA is a "whole-animal approach" used to assess welfare through the scrutiny of the animals' body language and using a number of descriptors such as *relaxed*, *anxious*, *content* or *frustrated*. These terms, given their emotional connotation, appear to have direct relevance to animal welfare by referring to the animals' own subjective experience (Wemelsfelder et al., 2001; Wemelsfelder and Farish, 2004; Wemelsfelder and Lawrence, 2001). Using principal component analysis, the number of variables is reduced to (usually) two main components, each comprising correlated, and to some degree overlapping, behavioural expressions. Interpretation of the main components involves the identification of the terms that best describe the anchor points at each end. The approach is the only measure in the WQ protocol that can capture positive emotional states (Keeling et al., 2013). For other farm animal species, QBA has been found to correlate in a biologically meaningful direction with physiological measures (Rutherford et al., 2012; Stockman et al., 2011; Wickham et al., 2015) and health measures (Phythian et al., 2016, des Roches et al., 2018). Painful conditions, like lameness, have for instance been seen to be associated with the QBA score in sheep, suggesting that compromised health had a wider deleterious effect on the sheep's emotional state (Phythian et al, 2016). Likewise, dairy cattle in the acute phase of E. coli mastitis were interpreted to experience a negative emotional state, as assessed with QBA (des Roches et al., 2018). With this in mind, QBA could potentially be used as a screening tool to identify flocks with health and pain issues that compromise the welfare to a degree where the animals' emotional state is affected. QBA is used in a few recently published studies on broiler welfare (Bassler et al., 2013; Buijs et al., 2017; de Jong et al., 2016; Federici et al., 2016). However, apart from one study in which no moderate or strong correlations were found between QBA scores and other measures in the WQ protocol (de Jong et al., 2016), there is a scarcity of information about how QBA relates to other animal-based welfare indicators in broilers.

Comprehensive animal welfare assessments, like the WQ protocol, usually also include behavioural indicators of affective states of particular relevance to animal welfare, such as fear of humans (Bassler et al., 2013; Forkman et al., 2007). Fear is one of the best-studied emotions in many farm animal species, and is generally expressed behaviourally as either active defence or avoidance, or passive avoidance (Forkman et al., 2007). In modern broiler production, there is little or no opportunity for day-to-day handling of the animals, so human-animal interactions are mainly limited to visual contact. In the WQ protocol for broiler chickens, fear of humans is assessed by the use of a touch test (Forkman et al., 2007). Previous papers have cast doubt about the validity of this test, as the chickens' avoidance of the assessor relies on their walking ability, which may be impaired in animals with poor leg health (de Jong et al., 2011; Vasdal et al., 2018). It is therefore of interest to investigate how this measure relates to other indicators of the chickens emotional state.

Poor leg health is a welfare issue of particular concern in industrialised broiler chicken production. Systematic recording of indicators of leg health, such as lameness and footpad dermatitis, are therefore included in the WQ protocol (WelfareQuality®, 2009). Poor leg health may be associated with both infectious and non-infectious factors and can cause commercial loss through increased mortality, culling and reduced performance (Butterworth, 1999; European Food Safety Authority, 2012; Scientific Committee on Animal Health and Animal Welfare, 2000). One common leg health issue is footpad dermatitis (FPD), which causes necrotic lesions and inflammatory processes on the plantar surface of the footpads in broilers. This may be painful and also cause lameness, and hence represents a valid and important indicator of broiler chicken welfare (Butterworth and Niebuhr, 2009). The condition is mainly caused by a variety of environmental factors, including wet litter (Shepherd and Fairchild, 2010). Lame birds (gait score ⩾3) have been shown to prefer food with analgesics, and increase their activity when given analgesics, which indicates that the observed lameness is associated with pain (McGeown et al., 1999; Danbury et al., 2000; Weeks et al., 2000). Lameness may also be associated with reduced activity in general, and less expression of positively motivated behaviours, which implies additionally compromised animal welfare (Sanotra et al., 2002; Weeks et al., 2000). Moreover, lame birds may have difficulties in reaching food and water (Butterworth et al., 2002; Sanotra et al., 2002), resulting in impaired growth and poor production results. Therefore, gait scoring is used to investigate severity of lameness in live birds (Kittelsen et al., 2017), and scoring of the macroscopic appearance of footpad lesions on farm or at the slaughterhouse is used to monitor welfare ante- or post mortem, respectively (Shepherd and Fairchild, 2010).

Mortality can be considered to be the animals' response to (or consequences of) risk factors (Jacobs et al., 2017) and therefore represents an important welfare indicator. On-farm mortality consists of both natural mortality (i.e. chickens found dead) and selective culling. High on-farm mortality can thus be an indicator of poor flock health, but may also reflect careful selection for culling by the stockperson.

Although QBA is included in the WQ protocol for poultry, there is little available knowledge about the use of this method in broiler chickens. Therefore, to gain more knowledge about QBA as a welfare assessment tool on broiler chicken farms, this study aims to describe the dimensionality of QBA, and to investigate the associations between QBA and selected animal-based welfare indicators from the Welfare Quality® protocol, i.e. lameness, foot pad dermatitis (FPD), fear of humans, and mortality.

2. Material and methods

2.1. Sampling

A total of 50 commercial broiler chicken farms were randomly recruited from the list of about 150 broiler producers delivering chickens (hybrid: Ross 308, mixed sex) to one major slaughter plant, located in the southeast of Norway (Nortura Hærland). The producers were contacted by phone a few weeks before the visit. Participation in the study was voluntary, however only one of the contacted farmers declined.

2.2. Farm visits and data collection

The farms were visited between January and March 2015. The entire Welfare Quality® protocol for broilers (WelfareQuality®, 2009) was performed, but only the selected measures of lameness, footpad dermatitis, fear of humans, and QBA will be described in this paper.

One of the authors (GV, an ethologist with comprehensive knowledge of broiler behaviour), had been trained in the theory and practice of the Welfare Quality® protocol by experienced WQ assessors, and performed all the farm visits. The visits were conducted between day 27 and 34 of the chickens' lives, on average (±SD) day 28.9 (±1.8). This was as close to slaughter as possible (average age of slaughter in Norway is 31 days), which is in accordance with recommendations in the Welfare Quality® protocol. Most broiler farms in Norway have only one house, therefore only one flock was assessed on each farm. None of the flocks were thinned.

The on-farm assessments on each farm were performed on the same day, and conducted in accordance with the methods and order described in the WQ assessment protocol (WelfareQuality®, 2009). All data was recorded using specialized software designed specifically for the WQ broiler protocol, on a personal digital assistant (PDA). The software was designed by H. van den Heuvel, Wageningen University and Research, Wageningen Livestock Research.

2.2.1 QBA

The assessments started with QBA. The observation of the flocks were done from different observation points in the broiler house, where the animals that could be seen well were observed for a total of 20 minutes, as described in the WQ protocol (WelfareQuality®, 2009). This was followed by scoring of the 22 behavioural expressions on visual analogue scales (VAS). Each 125 mm VAS ranged from 'Minimum', indicating that the behavioural expression is entirely absent in any of the animals observed, to 'Maximum', meaning the expressive quality is dominant across all observed animals. The behavioural expressions used were (in random order): *Active*, *Relaxed*, *Helpless*, *Comfortable*, *Calm*, *Content*, *Tense*, *Inquisitive*, *Friendly*, *Positively occupied*, *Scared*, *Drowsy*, *Fearful*, *Agitated*, *Confident*, *Depressed*, *Unsure*, *Energetic*, *Frustrated*, *Bored*, *Playful*, *Nervous*, and *Distressed*.

2.2.2 Gait score and foot pad dermatitis (FPD)

After the QBA, the assessor then gait scored 150 randomly selected birds from at least five locations representing different areas of the house, such as near the walls and the center. About 30 birds were carefully fenced in at each location, using a mobile catching pen that could fence a group of animals without much disturbance. Each bird was then individually encouraged to walk out of the pen to be gait scored. To avoid affecting the birds' gait, no birds were handled or picked up prior to gait scoring. A six point rating scale was used, ranging from 0 (normal, dexterous and agile) to 5 (incapable of walking) (Kestin et al., 1992). After the gait scoring, a total of 100 random birds from five new locations (around 20 birds in each location) were carefully fenced in and scored for footpad dermatitis by visual inspection of their footpads. FPD was scored from 0 (no footpad lesion) to 4 (severe lesion, large area injured).

2.2.3 Touch test

In the touch test, the assessor approached a group of at least three birds, squatted for 10 seconds and then recorded the number of birds at an arm's length (i.e. within 1 meter of the observer), and the number of birds actually touched. Every attempt to approach a group of birds was considered a trial, even if all birds from the group withdrew from the approaching or squatting assessor. Twenty-one trials were conducted at several different locations around the house, to avoid repeated scoring of the same birds. If no animals were within an arm's length within the first 12 trials, the touch test was terminated.

2.2.4 Mortality

Shortly after each flock was slaughtered, production records including mortality rate, growth rate and rejection causes, were collected from the slaughter house (Nortura Hærland). Only the mortality data are presented in this study. Total mortality is the number of birds delivered to the slaughterhouse subtracted from number of birds delivered to the farmer from the hatchery, and the flock mortality rate is calculated as the percentage of dead birds. The farmers were given information about the use of these data at the time of the farm visits.

2.3 Data management and statistical analyses

All statistical analyses were performed in Stata SE/14.2 (StataCorp, College Station, TX, USA).

2.3.1 Calculation of component scores for qualitative behavioural assessments

QBA scores (i.e. the distance between the *Minimum* point on the visual analogue scale, to the mark made by the observer, providing a value between 0 and 125), as registered on the hand-held device were exported to Microsoft Office Excel® 2010, and subsequently transferred to Stata SE/14.2. Principal component analysis (PCA) was conducted using a correlation matrix (no rotation), retaining the two components that explained most of the variance in the data. Two new variables, PC1 and PC2, representing the scores for the two main components were generated.

2.3.2 Calculation of gait score

Gait score, which was assessed for 150 animals in each flock, was calculated by multiplying the number of animals with score 0 by 0, the number of animals with score 1 by 1, and so on up to score 5:

$\sum=((n0*0)+(n1*1)+(n2*2)+(n3*3)+(n4*4)+(n5*5))$. The total flock score could theoretically range between 0 (all 150 animals receive score 0) and 750 (all 150 animals receive score 5).

2.3.3 Calculation of foot pad dermatitis score

Footpad dermatitis scores from on-farm assessments were calculated by multiplying the number of animals (of the 100 examined) with score 0 with 0, the number of animals with score 1 or 2 by 1, and animals with score 3 or 4 by 2 (i.e. $\sum = (n0*0) + ((n1 + n2)*1) + ((n3 + n4)*2)$). The total flock score could theoretically range between 0 (all 100 animals receive FPD score 0) and 200 (all 100 animals receive FPD score 3 or 4).

2.3.4 Calculation of touch test score

For the touch test, calculations were performed in accordance with the description in the WQ protocol (WelfareQuality®, 2009):

The theoretical number of bird that should be within arm's reach of the observer if the birds were evenly spread in the barn is calculated from stocking density. This theoretical number is equal to the stocking density (expressed in birds/m^2) multiplied with $\pi/2$ (we divide by two the exact surface of a circle which radius in 1 m, to cover for the space taken by the observer). The number of birds that are within arm's reach of the observer (i.e. within 1 m) was compared to that theoretical number of birds. An index representing the % birds within 1 m is calculated: I = 100 × (number of birds within arm's reach/theoretical number of birds). The index is turned into a score according to the following spline functions:

When $I \leqslant 20$ then Score = $24.631 + (8.9944 \times I) - (0.32423 \times I^2) + (0.0031378 \times I^3)$

When $I \geqslant 20$ then Score = $95.660 + (0.46453 \times I) - (0.014127 \times I^2) + (8.7479 \times I^3)$

These calculations resulted in a touch test score for each of the 50 flocks. The touch test score can theoretically range from 24.6 (no animals touched) to 100 (all animals that theoretically can be touched, are touched). Thus, an increased touch test score is meant to indicate a reduced fear of humans and an improved human–animal relationship.

2.3.5 Regression analyses

Regression analysis was used to assess the associations between QBA and other welfare indicators. Footpad score, lameness score, touch test score, and mortality were entered as dependent variables in the regression analyses, with the two main components from the principal component analysis of QBA as the independent variables.

Because of strongly skewed distributions of several of the outcome variables, and non-linear associations with QBA as assessed by screening with linear regression and graphical methods, we needed to transform the variables prior to running the regression analyses. This was done by categorising them and running logistic or ordinal logistic regression analysis (see details for each variable below). Ordinal logistic regression analysis is based on a single equation with only one coefficient for each independent variable, and thus assumes proportional odds. To test this assumption, two tests were performed on each model; the Brant Test of Parallel Regression Assumption (Brant, 1990) and an approximate likelihood ratio-test (Wolfe and Gould, 1998).

The lameness score had a certain degree of right skew. Log transformation did not resolve this completely. To avoid violation of the major assumptions of linear regression we therefore categorised this variable into three equally large categories: low gait score (n=17), medium gait score (n=18) and high gait score (n=15). The association between this variable and QBA was therefore assessed with ordinal logistic regression.

The footpad score was right skewed, and three categories were created, which was the maximum number we considered feasible for the data, representing low (n=18), medium (n=16) and high (n=16) footpad score. Thus, for this outcome we also ran ordinal logistic regression.

The touch test score had the strongest right skew, and it was not considered feasible to divide the data into more than two categories. The variable was therefore dichotomized into flocks in which no birds were possible to touch (n=20, score=24.6) and flocks were at least some birds allowed the observer to touch them (n=30, score>24.6). The dichotomized touch test score was thus tested in a logistic regression model.

Mortality was also somewhat right skewed, and screening with linear regression and the “lintrend” command in Stata suggested a non-linear relationship with PC1. The variable was categorised into three quantiles of equal size: low (n=17), medium (n=17) and high (n=16) mortality, and associations with QBA was thus investigated using ordinal logistic regression analyses.

3. Results

3.1. General farm results

The mean (±SD) flock size in the 50 visited farms was 17391 (±6080) chickens, and ranged from 3900 to 28950 birds. The chickens’ mean (±SD) age was 28.9 days (±1.8) at the time of visit, and ranged from 27 to 34 days. The mean animal density was 17.4 kg/m^2, with a range of 22.2 to 33.18 kg/m^2. Other descriptive flock statistics have been presented elsewhere (Vasdal et al., 2018).

3.2 QBA

The principal component analysis of the QBA data revealed two main dimensions, explaining 48.3% and 22.1% of the variance respectively (70.4% overall). The scatterplot in Figure 1 illustrates the component loadings of each behavioural term across the two principal components. The first component ranged from *relaxed*, *calm* and *drowsy* to, *agitated*, *fearful*, *tense* and *nervous*, but also terms with a positive connotation, such as *energetic*, *positively occupied* and *playful*, loaded highly on this component. This component was labelled *arousal*. The second component ranged from *depressed*, *frustrated* and *distressed* to *friendly*, *content*, *comfortable* and *confident*, and was labelled *mood.*

Figure 1 here.

3.3. Selected animal based measures

Mean lameness score was 259.4, of a theoretical maximum of 750. The majority of the chickens had a lameness score of 1 (44%) or 2 (34%). For the distribution across the gait score categories, see Vasdal et al. (2018). Mean footpad dermatitis score was 15.5, indicating that most farms had a low prevalence of footpad dermatitis. Touch test scores ranged from 24.6 to 99.8, with a mean of 45.1, in other words, with a strong right skew (Table 1).

Table 1. Mean (±SD) and range of scores for selected animal-based welfare measures from 50 Norwegian broiler chicken flocks.

Animal-based welfare indicator	**Mean (±SD)**	**Range**
Gait score	259.4 (±52.0)	186 – 439
Foot pad dermatitis (FPD) score	15.5 (±22.4)	0 – 111
Touch test (TT) score	45.1 (±31.4)	24.6 - 99.9
Mortality rate	2.2 (±0.8)	1.1 - 5.4

3.4 Regression analyses

3.4.1 Lameness as outcome

Ordinal logistic regression revealed that none of the two components of QBA were significantly associated with the categorized gait score variable, with p = .20 for PC1 (*arousal*) and p = .14 for PC2 (*mood*). The Brant test and the approximate likelihood ratio test were both non-significant, indicating that the assumption of parallel regression was not violated.

3.4.2 Footpad score as outcome

Ordinal logistic regression revealed that none of the two components of QBA were significantly associated with the categorized footpad score variable, with p = .26 for PC1 (*arousal*) and p = .61 for PC2 (*mood*). The Brant test and the approximate likelihood ratio test were both non-significant, indicating that the assumption of parallel regression was not violated.

3.4.3 Touch test as outcome

Logistic regression analysis revealed that both PC1 (*arousal*) and PC2 (*mood*) were significantly and negatively associated with the dichotomised touch test variable, with 26% of the variance in the touch test explained by the model (pseudo-R^2 = 0.26). Flocks with a high score on either *arousal* or *mood* were less likely to have birds that accepted being touched by the observer (*arousal*: OR = 0.70, p = .004; *mood*: OR = 0.64, p = .020).

Figure 2 here.

3.4.4 Mortality as outcome

Ordinal logistic regression revealed that *arousal* (PC1) was significantly and negatively associated with the categorized variable for mortality, i.e. flocks with a higher arousal score were less likely to be scored in a higher mortality category at the time of slaughter (OR = 0.81, p = .02). The mood score (PC2) was unrelated (OR = 1.00, p = .99) with mortality. The univariable model with PC1 explained 6% of the variance in mortality (pseudo-R^2 = 0.06). The Brant test and the approximate likelihood ratio test were both non-significant, indicating that the assumption of parallel regression was not violated.

Figure 3 here.

Table 2. Odds ratios (SE), p-values and 95 % confidence intervals from regression models with selected animal-based welfare measures as dependent variables and the two main components of QBA (PC1 and PC2) as independent variables. The explained variance of the models is reported as pseudo R^2. Significant p-values ($p < .05$) in bold.

Dependent variable	Independent variable	OR	SE	*P*	95% C.I.	Pseudo R^2
Gait score[a]	PC1	0.91	0.08	0.20	-0.25 – 0.05	0.04
	PC2	0.84	0.12	0.14	- 0.40 – 0.06	
FPD score [a]	PC1	1.09	0.08	0.26	-0.07 – 0.25	0.01
	PC2	0.94	0.11	0.61	- 0.28 – 0.16	
Touch test score[b]	PC1	0.70	0.09	< **0.01**	0.54 – 0.89	0.26
	PC2	0.64	0.12	**0.02**	0.44 – 0.93	
Mortality rate [a]	PC1	0.81	0.71	**0.02**	0.68 – 0.96	0.06
	PC2	1.00	0.12	1.0	0.80 – 1.26	

[a] Ordinal logistic regression

[b]Logistic regression

4. Discussion

The present study seeks to contribute to the evaluation of QBA as a method of assessing broiler welfare, by investigating its dimensionality and whether scores are associated with other welfare outcomes considered to be of importance in broiler chicken production. Health conditions that are known to have a detrimental effect on the animals' welfare, could be expected to affect measures of their emotional state.

4.1 The dimensionality of qualitative behavioural assessments in broiler chickens

Principal component analysis (PCA) of data from the qualitative behavioural assessments in 50 broiler flocks revealed a dimensionality that can be recognised from QBA in other species (Brscic et al., 2009 [veal calves]; Duijvesteijn et al., 2014 [pigs]; Grosso et al., 2016 [goats]; Minero et al., 2016 [donkeys]). In these and other studies mood has usually been identified as the first component, whereas in our study the arousal-component explains somewhat more of the variance than the mood component. The first component ranged from terms associated with low arousal (*relaxed*, *calm* and *drowsy*) to terms associated with high arousal, with both positive and negative emotional connotations (e.g. *fearful*, *nervous*, *energetic,* and *playful*). The second component (*mood*) ranged from negatively connoted terms such as *depressed* and *distressed*, to positively connoted terms such as *friendly* and *confident.*

4.2 Lameness

None of the two components of QBA were significantly associated with the categorized gait score variable. Studies have found that between 13% to 30% of broilers worldwide have an impaired gait (i.e. gait score 3, 4 or 5) (e.g. Bassler et al., 2013; Kittelsen et al., 2017; Louton et al., 2018). In the 50 farms visited in this study, 19 % of the birds had a gait score ≥3, including 2.4 % with score 4 and 0.5 % with score 5 (presented in Vasdal et al., 2018). Federici et al. (2016) found lameness to be a considerable welfare problem in Brazilian broiler flocks, assessed using Welfare Quality® measures. The observed median percentage of severe lameness (scores 4 and 5) in their study were on average 14 % (range 4 % - 27 %), and was hence considerably higher than the 2.9 % prevalence of severe lameness found in our data

(Vasdal et al., 2018). Mean flock age at visit was 40 days (range 35 – 44) in the Brazilian study, compared to 29 days in our study. It is likely that this may have affected the discrepancy in results, as lameness has been found to increase in severity as the chickens' age increases (e.g. Silvera et al., 2016). Silvera et al. (2016) found that gait score was unaffected by increasing human contact (i.e. improving the human – animal relationship). The high median score Federici et al (2016) obtained in the touch test could be confounded by the high percentage of severe lameness found in the selected flocks, making the birds less able or willing to move away from the observer. As presented in a previous paper (Vasdal et al., 2018), there was an association between the touch test score and lameness in the data from the 50 farms used in this study. This implies that the touch test may be confounded by the chickens' reduced ability to walk. In line with this, Louton et al. (2018) found that a gait score ≥3 was associated with lower weights, suggesting that the chickens' ability to walk was impaired due to the lameness, hence reducing their feed intake.

4.3 Footpad dermatitis score

The footpad dermatitis score ranged from 0 to 111, with an average score of 17.82, indicating that most farms had a low prevalence of footpad dermatitis. Neither of the two components of QBA were significantly associated with the categorized footpad score variable. A consistently low prevalence (i.e. a homogenic population) makes it more difficult to prove associations statistically, hence requiring larger sample sizes (Houe et al., 2004). Louton et al. (2018) found no association between FPD and other health or management-related welfare indicators in their study from the USA, but they reported a worsening of the FPD scores over time. However, the assessments in their study continued until fattening day 39, thus the animals were about ten days older than the chickens were at the time of the assessment in this study.

4.4 Touch test

The level of fear of humans, as measured by the touch test, revealed differences between flocks with regards to numbers of animals possible to touch. Flock scores ranged from 24.6 (no animals touched) to 99.88 (186 animals touched), with an average of 45.13 (corresponding to approximately 29 animals touched). In 30 of the 50 flocks, the observer was not able to touch any birds. Silvera et al. (2016) found that the proportion of animals touched in their experiment was significantly increased following additional human contact. Variations between flocks in our study could be therefore potentially be related to the quantity or quality of the farmers' interactions with the broilers. However, a thorough assessment of the human – animal relationship was beyond the scope of this study. In the investigation of associations between QBA and the touch test, we found that flocks with higher arousal (PC1) and/or mood (PC2) scores were less likely to have birds that accepted being touched by the observer. Bassler et al. (2013) used QBA to assess 89 broiler flocks and found that the same flocks showed both agitated/fearful and inquisitive/playful patterns of expression. They suggested that the seemingly contradictory outcomes may be two sides of the same coin: both expressing greater responsivity, or in other words, greater arousal or liveliness in interaction with the environment (Bassler et al., 2013). Similarly, both higher arousal and mood scores may indicate greater liveliness and responsivity in birds that did not accept to be touched by the observer in our study, confounding the indicator of fear. It has already been suggested that the touch test is confounded by impaired walking ability (de Jong et al., 2011; Vasdal et al., 2018). Moreover, Silvera et al. (2016) reported that they were unable to use the touch test in the last week of the rearing period due to the crowded conditions, restricting the birds' ability to move away from the observer. The test is therefore a suboptimal method of assessing fear of humans in broilers, at least at the end of the rearing period. In our study, the higher arousal or mood scores associated with fewer animals touched, may also indicate a better leg health and walking ability, rather than fear of the observer. In some of the observed flocks, the escaping birds would return to the squatting assessor after a few moments. This suggests that these birds were motivated to approach the assessor after they initially fled (Vasdal et al., 2018). de Jong et al. (2011) also observed that flocks scored as being fearful for humans in the touch test, appeared not to be fearful in the opinion of the observers. Chickens can copy the behaviour of their companions, which may lead to all individuals showing simultaneous flight or escape behaviour (Nicol, 2015). This acts to increase the degree of synchrony within the group, and may have had an effect on the initial responses to the touch test. Moreover, stocking density (a lower density makes it easier for the birds to flee) and greater light intensity (making the birds more responsive to their surroundings) may have an influence on this measure (Tuyttens et al., 2015). All this calls for validation studies and/or development of alternative methods of testing fear in broiler chickens. Fear is one of many aspects assessed within the QBA method, but due to the integrative nature of the method, it does not provide information on this emotion specifically.

4.5 Mortality

In our study, we found that flocks with a higher arousal score as measured by the first component of QBA (PC1) were less likely to be in a higher mortality category. Greater ability to express arousal may be an indication of better flock health (e.g. good leg health and walking ability) and less apathy. However, the percentage of variation in mortality explained by PC1 was low (6 %), meaning that the potential for using QBA to predict an increased risk of higher mortality is very low. Mortality is comprised of controlled and uncontrolled events, in which controlled mortality are the birds culled humanely by the producer, whereas uncontrolled mortality consists of the birds left to die, which inevitably will be associated with prolonged suffering for some individuals. Careful selective culling can result in a reduction of the percentage of animals dead on arrival (%DOA) at the slaughter plant (Jacobs et al., 2017), as fewer animals of poor health are subjected to the stress of transport. The proportion of controlled mortality should ideally be maximized compared to uncontrolled mortality, for animal welfare reasons (Butterworth and Niebuhr, 2009). Although culling should not be needed in an optimal situation, high proportions of culling can reflect that measures are in place to prevent animal suffering if sick or injured (European Food Safety Authority, 2010). The mortality rate in our study does not distinguish between culling and natural deaths (the data for culling rate was incomplete). However, it is reasonable to assume that the majority of animals that were culled would have died naturally before the time of slaughter if they were not culled, so the total mortality rate can be considered measure that provides information about the overall flock health.

4.6 Limitations

A limited sample size may have reduced our possibility of finding statistically significant associations between QBA and the other animal-based welfare indicators. Andreasen et al. (2013) failed to find meaningful relationships between QBA scores and other Welfare Quality® measures in dairy cattle, and suggested that the spread between the farms in their study was too small to robustly anchor an effective qualitative welfare scale to correspond with WQ outcomes. This may have reduced the possibility of detecting associations in this study too. For several parameters, our data were homogenous, i.e. with little variation between the flocks observed, reducing the possibility to detect associations. This may be an inherent issue in an industry where the production system and management in general is highly standardised, as is the case in broiler production. This was also the case when Buijs et al. (2017) tested the sensitivity of the WQ broiler protocol to detect differences between intensively reared flocks, where they experienced that the observed values for health parameters often where extreme (either very high or very low). For FPD and lameness, the scores were low, and the observed range for measures of appropriate behaviour were very narrow, except for QBA (Buijs et al., 2017). This wider variation in QBA scores may thus make it more suitable for detecting variation among farms, given that its validity is established.
The QBA-method has not been validated for broiler chickens (de Jong et al., 2014; Wemelsfelder et al., 2009), and observers need to have sufficient knowledge of broiler chickens and their behaviour for a reliable and valid scoring. In our study, the observer had a comprehensive knowledge of behaviour in chickens. It may be easier to score QBA for larger animals kept in smaller groups, allowing a better observation of postures, facial expressions and vocalisations, as compared to broilers housed in groups of several thousands (de Jong et al., 2014).

4.7 Conclusions

In this study, a negative association was found between both components of QBA and the results of the touch test, which is designed to measure the chickens' fear of humans. In other words, flocks with higher scores on both *arousal* and *mood*, were less likely to have any chickens that were possible to touch by the assessor. This raises further questions about the validity of the touch test as a measure of animal welfare. For the other selected animal-based measures, there were no associations with QBA, except for mortality. In accordance with the findings in Andreasen et al. (2013), the current study does not support the idea that QBA can be used as a stand-alone on-farm welfare assessment tool, capable of predicting the other important welfare outcomes from the WQ protocol. However, this method may give valuable supplementary information, but must first be thoroughly validated as a welfare assessment tool for broilers.

Acknowledgements

This work was supported by the Norwegian Research Council (NFR grant no 234191). The authors would like to thank Henk Gunnink for valuable training of our observer in the use of the Welfare Quality® protocol. The authors would also like to thank all participating farmers for allowing us into their farms, and Ms. Anne Mette Dagrød and Ms. Hilde Bryhn (both Nortura) for efficiently providing us with production data from the visited flocks.

References

Andreasen, S.N., Wemelsfelder, F., Sandøe, P., Forkman, B., 2013. The correlation of Qualitative Behavior Assessments with Welfare Quality® protocol outcomes in on-farm welfare assessment of dairy cattle. Appl. Anim. Behav. Sci. 143, 9–17. doi:http://dx.doi.org/10.1016/j.applanim.2012.11.013

Bassler, A., Arnould, C., Butterworth, A., Colin, L., de Jong, I., Ferrante, V., Ferrari, P., Haslam, S., Wemelsfelder, F., Blokhuis, H.J., 2013. Potential risk factors associated with contact dermatitis, lameness, negative emotional state, and fear of humans in broiler chickens flocks. Poult. Sci. 92, 2811–2826.

Boissy, A., Manteuffel, G., Jensen, M.B., Moe, R., Spruijt, B., Keeling, L.J., Winckler, C., Forkman, B., Dimitrov, I., Langbein, J., Bakken, M., Veissier, I., Aubert, A., 2007. Assessment of positive emotions in animals to improve their welfare. Physiol. Behav. 92, 375–397.

Brant, R., 1990. Assessing proportionality in the proportional odds model for ordinal logistic regression. Biometrics 46, 1171–1178.

Brscic, M., Wemelsfelder, F., Tessitore, E., Gottardo, F., Cozzi, G., Van Reenen, C.G., 2009. Welfare assessment: Correlations and integration between a Qualitative Behavioural Assessment and a clinical/health protocol applied in veal calves farms. Ital. J. Anim. Sci. 8 (Suppl., 601–603.

Buijs, S., Ampe, B., Tuyttens, F.A.M., 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.

Butterworth, A., 1999. Infectious components of broiler lameness: A review. Worlds. Poult. Sci. J. 55, 327–352. doi:10.1079/WPS19990024

Butterworth, A., Niebuhr, K., 2009. Measures of poultry health status, in: Forkman, B., Keeling, L. (Eds.), Welfare Quality Reports No. 9. SLU Service/Reproenheten, Uppsala, Sweden, pp. 39–65.

Butterworth, A., Weeks, C., Crea, P., Kestin, S., 2002. Dehydration and lameness in a broiler flock. Anim. Welf. 11, 89–94.

Danbury T.C., Weeks C.A., Chambers J.P., Waterman-Pearson A.E., Kestin S.C., 2000. Self-selection of the analgesic drug carprofen by lame broiler chickens. Vet. Rec. 146, 307–311.

de Jong, I., Gunnink, H., Hindle, V., 2014. Implementation of the Welfare Quality® broiler assessment protocol – final report. Overview of outcome-based measurement of broiler welfare and a general discussion on the Welfare Quality® broiler assessment protocol. Wageningen UR Livestock Research Report 833. 58 blz., Wageningen, The Netherlands.

de Jong, I., Hindle, V., Butterworth, A., Engel, B., Gunnink, H., Perez Moya, T., Tuyttens, F.A.M., Van Reenen, C.G., 2016. Simplifying the Welfare Quality assessment protocol for broiler chicken welfare. Animal 10, 117–127.

de Jong, I., Reimert, H., Vanderhasselt, R., Gerritzen, M., Gunnink, H., van Harn, J., Hindle, V., Lourens, A., 2011. Development of methods to monitor foot pad lesions in broiler chickens. Wageningen UR Livestock Research Report 463. Wageningen, The Netherlands.

des Roches, A.B., Lussert, A., Faure, M., Herry, V., Rainard, P., Durand, D., Wemelsfelder, F., Foucras, G., 2018. Dairy cows under experimentally-induced Escherichia coli mastitis show negative emotional states assessed through Qualitative Behaviour Assessment, Appl. Anim. Behav. Sci. 206, 1-11.

Duijvesteijn, N., Benard, M., Reimert, I., Camerlink, I., 2014. Same pig, different conclusions: Stakeholders differ in qualitative behaviour assessment. J. Agric. Environ. Ethics 27, 6.

European Food Safety Authority (EFSA), 2012. Scientific Opinion on the use of animal-based measures to assess welfare of broilers. Parma, Italy.

European Food Safety Authority (EFSA), 2010. Scientific Opinion on the influence of genetic parameters on the welfare and the resistance to stress of commercial broilers. Parma, Italy.

Federici, J., Vanderhasselt, R., Sans, E., Tuyttens, F.A.M., Souza, A., Molento, C., 2016. Assessment of broiler chicken welfare in Southern Brazil. Brazilian J. Poult. Sci. 18, 133–140.

Forkman, B., Boissy, A., Meunier-Salaün, M.-C., Canali, E., Jones, R.B., 2007. A critical review of fear tests used on cattle, pigs, sheep poultry and horses. Physiol. Behav. 92, 340–374.

Green, T., Mellor, D.J., 2011. Extending ideas about animal welfare assessment to include "quality of life" and related concepts. N. Z. Vet. J. 59, 263–271. doi:10.1080/00480169.2011.610283

Grosso, L., Battini, M., Wemelsfelder, F., Barbieri, S., Minero, M., Dalla, E., Silvana, C., 2016. On-farm qualitative behaviour assessment of dairy goats in different housing conditions. Appl. Anim. Behav. Sci. In Press. doi:doi:10.1016/j.applanim.2016.04.013

Houe, H., Kjær Ersboll, A., Toft, N., 2004. Introduction to veterinary epidemiology. Biofolia, Fredricksberg, Denmark.

Jacobs, L., Delezie, E., Duchateau, L., Goethals, K., Tuyttens, F.A.M., 2017. Broiler chickens dead on arrival: associated risk factors and welfare indicators. Poult. Sci. 96, 259–265.

Keeling, L., Evans, A., Forkman, B., Kjærnes, U., 2013. "Welfare Quality principles and criteria." Wageningen, The Netherlands.

Kestin, S., Knowles, T., Tinch, A., Gregory, N.G., 1992. Prevalence of leg weakness in broiler chickens and its relationship with genotype. Vet. Rec. 131, 190–194.

Kittelsen, K., David, B., Moe, R., Poulsen, H., Young, J., Granquist, E., 2017. Associations between gait score, production data, abattoir registrations and post mortem tibia measurements in Norwegian broiler chickens. Poult. Sci. 96, 1033–1040.

Louton, H., Bergmann, S., Reese, S., Erhard, M., Bachmeier, J., Rösler, B., Rauch, E., 2018. Animal- and management-based welfare indicators for a conventional broiler strain in 2 barn types (Louisiana barn and closed barn). Poult. Sci. 97, 2754–2767.

McGeown D.T., Danbury T., Waterman-Pearson A., Kestin S., 1999. Effect of carprofen on lameness in broiler chickens. Vet. Rec. 144, 668–671.

Mellor, D.J., 2016. Opinion: Updating animal welfare thinking: Moving beyond the "Five Freedoms" towards "A life worth living." Animals 6, 21.

Minero, M., Dalla Costa, E., Dai, F., Murray, L., Canali, E., Wemelsfelder, F., 2016. Use of qualitative behaviour assessment as an indicator of welfare in donkeys. Appl. Anim. Behav. Sci. 174, 147–153.

Nicol, C.J., 2015. The behavioural biology of chickens. CABI, Wallingford, UK. doi:10.1079/9781780642499.0000

Phythian, C.J., Michaelopoulou, E., Cripps, P.J., Duncan, J.S., Wemelsfelder, F., 2016. On-farm qualitative behaviour assessment in sheep: Repeated measurements across time, and association with physical indicators of flock health and welfare. Appl. Anim. Behav. Sci. 175, 23–31.

Rutherford, K.M.D., Donald, R.D., Lawrence, A.B., Wemelsfelder, F., 2012. Qualitative behavioural assessment of emotionality in pigs. Appl. Anim. Behav. Sci. 139, 218–224. doi:http://dx.doi.org/10.1016/j.applanim.2012.04.004

Sanotra, G., Lund, J., Vestergaard, K., 2002. Influence of light-dark schedules and stocking density on behaviour, risk of leg problems and occurrence of chronic fear in broilers. Br. Poult. Sci. 43, 344–354.

Scientific Committee on Animal Health and Animal Welfare (SCAHAW), 2000. The welfare of chickens kept for meat production (broilers) (SCAHAW Report No. SANCO.B3/AH/R15/2000). Brussels, Belgium.

Shepherd, E., Fairchild, B., 2010. Footpad dermatitis in poultry. Poult. Sci. 89, 2043–2051.

Silvera, A., Wallenbeck, A., Butterworth, A., Blokhuis, H.J., 2016. Modification of the human–broiler relationship and its potential effects on production. Acta Agric. Scand. Sect. A - Anim. Sci. 66, 161–167.

Stockman, C.A., Collins, T., Barnes, A.L., Miller, D., Wickham, S.L., Beatty, D.T., Blache, D., Wemelsfelder, F., Fleming, P.A., 2011. Qualitative behavioural assessment and quantitative physiological measurement of cattle naïve and habituated to road transport. Anim. Prod. Sci. 51, 240–249.

Tuyttens, F.A.M., Federici, J., Vanderhasselt, R., Goethals, K., Duchateau, L., Sans, E., Molento, C., 2015. Assessment of welfare of Brazilian and Belgian broiler flocks using the Welfare Quality protocol. Poult. Sci. 94, 1758–1766.

Vasdal, G., Moe, R., de Jong, I., Granquist, E., 2018. The relationship between measures of fear of humans and lameness in broiler chicken flocks. Animal 12, 334–339.

Weeks, C., Danbury, T., Davies, H., Hunt, P., Kestin, S., 2000. The behaviour of broiler chickens and its modification by lameness. Appl. Anim. Behav. Sci. 67, 111–125.

WelfareQuality®, 2009. WelfareQuality® assessment protocol for poultry (broilers, laying hens). The WelfareQuality® Consortium, Lelystad, The Netherlands.

Wemelsfelder, F., Farish, M., 2004. Qualitative categories for the interpretation of sheep welfare: a review. Anim. Welf. 13, 261–268.

Wemelsfelder, F., Hunter, E.A., Mendl, M.T., Lawrence, A.B., 2001. Assessing the “whole animal”: a free choice profiling approach. Anim. Behav. 62, 209–220.

Wemelsfelder, F., Knierim, U., Schulze Westerath, H., Lentfer, T., Staack, M., Sandilands, V., 2009. Qualitative behaviour assessment, in: Forkman, B., Keeling, L. (Eds.), Assessment of Animal Welfare Measures for Layers and Broilers, Welfare Quality Reports No. 9. University of Cardiff, Cardiff, UK, pp. 113–119.

Wemelsfelder, F., Lawrence, A.B., 2001. Qualitative assessment of animal behaviour as an on-farm welfare-monitoring tool. Acta Agric. Scand. Sect. A - Anim. Sci. 51, 21–25.

Wickham, S.L., Collins, T., Barnes, A.L., Miller, D.W., Beatty, D.T., Stockman, C., Blache, D., Wemelsfelder, F., Fleming, P.A., 2015. Validating the use of qualitative behavioral assessment as a measure of welfare og sheep during transport. J. Appl. Anim. Welf. Sci. 18, 269–286.

Wolfe, R., Gould, W., 1998. An approximate likelihood-ratio test for ordinal response models. Stata Tech. Bull. 42, 24–27.

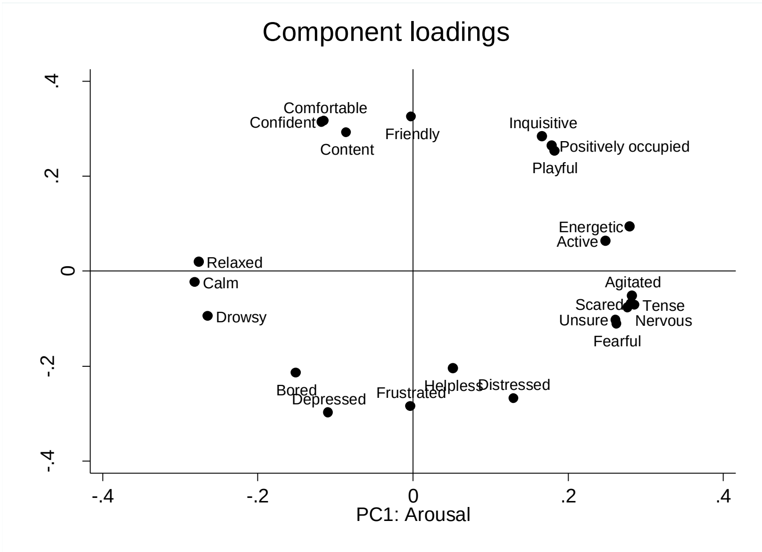
Figure 1: Loading plot illustrating the component loadings of each behavioural term across the two main components, a*rousal* (PC1) and m*ood* (PC2). These components account for 70.4% of the variance from the principal component analysis of QBA data from the 50 Norwegian broiler flocks.

Figure 2: Box plots illustrating the different distributions (median, interquartile range and range) of PC1 (arousal) and PC2 (mood) among the broiler flocks within the two categories of the touch test (TT) used in the regression analysis (0 = no broiler chickens touched and 1 = some broiler chickens touched).

Figure 3: Box plot illustrating the different distributions (median, interquartile range and range) of the first component (PC1 *arousal*) of the qualitative behavioural assessments in 50 Norwegian broiler flocks within different mortality categories (1 = low, 2 = medium, and 3= high).

619 Figure 1:

620



621

622

Figure 2:

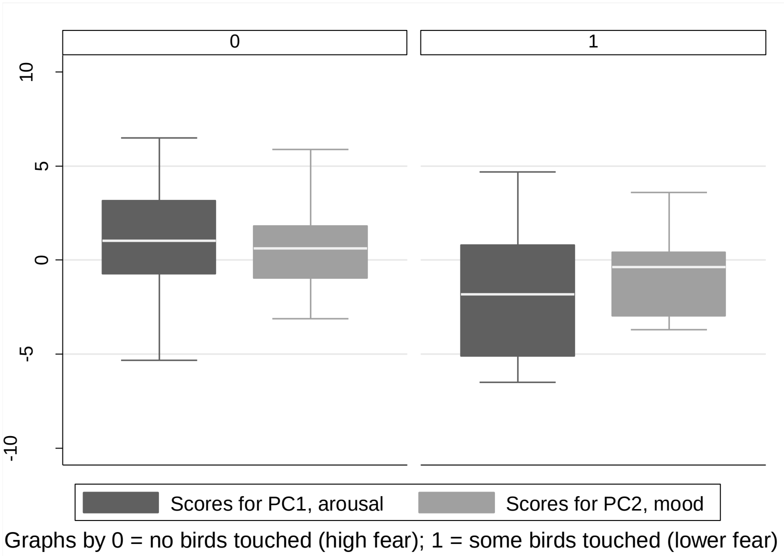
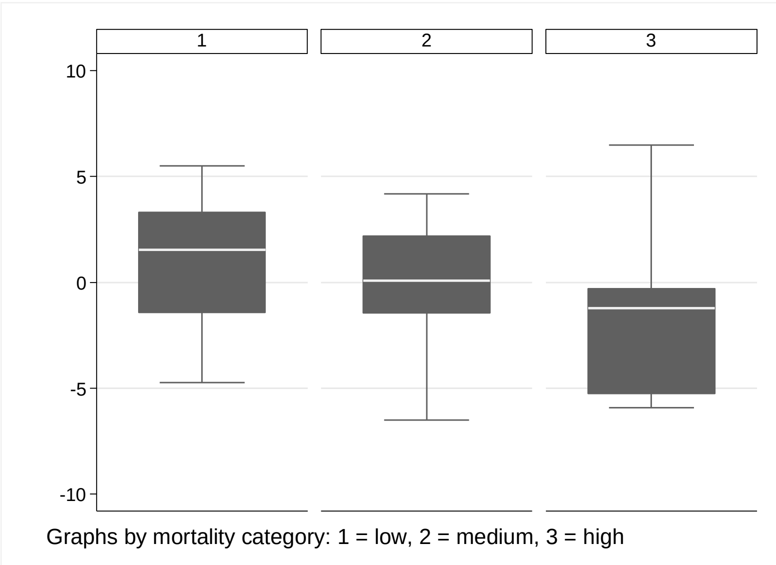

0
1
10
5
0
-5
-10
Scores for PC1, arousal
Scores for PC2, mood
Graphs by 0 = no birds touched (high fear); 1 = some birds touched (lower fear)

Figure 3:


1
2
3
10
5
0
-5
-10
Graphs by mortality category: 1 = low, 2 = medium, 3 = high