Animal Welfare 2020, 29: 185-196
ISSN 0962-7286
doi: 10.7120/09627286.29.2.185

Validation of histological and visual scoring systems for foot-pad dermatitis in broiler chickens

A Piller[†]*, S Bergmann*[†]*, A Schwarzer*[†]*, M Erhard*[†]*, J Stracke*[‡]*, B Spindler*[‡]*, N Kemper*[‡]*, P Schmidt*[§]*, J Bachmeier*[#]*, B Schade*[¶]*, B Boehm*[¶]*, E Kappe*[¶] *and H Louton*[*†]

[†] Department of Veterinary Sciences, Chair of Animal Welfare, Animal Behaviour, Animal Hygiene and Animal Husbandry, Faculty of Veterinary Medicine, LMU Munich, Veterinärstraße 13/R, 80539 Munich, Germany

[‡] Institute for Animal Hygiene, Animal Welfare and Farm Animal Behaviour, University of Veterinary Medicine Hannover, Foundation, Bischofsholer Damm 15, 30173 Hannover, Germany

[§] Paul Schmidt, Statistical Consulting for Science and Research, Große Seestr 8, 13086 Berlin, Germany

[#] Veterinary Practice, Grüner Weg 19, 94315 Straubing, Germany

[¶] Bavarian Animal Health Service, Department of Pathology, Senator-Gerauer-Str 23, 85586 Poing, Germany

* Contact for correspondence: h.louton@lmu.de

Abstract

In this study, an appropriate visual scoring system for foot-pad dermatitis was validated, considering the histologically measured depth of the inflammation zone and the histopathological grade (no lesion, mild lesion, ulcer). The aim being to evaluate whether the visual, macroscopic scoring of foot-pad dermatitis can represent the histological, microscopic findings. Two hundred Ross 308 broiler chicken feet (birds aged 39–42 fattening days) were collected at a slaughterhouse and scored macroscopically according to a modified version of the Welfare Quality® Assessment Protocol for Poultry. Afterwards, 200 histological slides (one per foot) were prepared, the extent of the inflammation measured and all slides scored by veterinarian pathologists using Michel et al*'s modified scheme. The statistical relationship between microscopic and macroscopic score and depth of inflammation were estimated via regression models. Increasing macroscopic score was found to be linked with an increase in microscopic score and the depth of inflammation. In particular, feet without lesions and feet with ulcers were identifiable using the macroscopic score. Macroscopic scoring of foot-pad dermatitis can mirror histological findings once certain limitations are taken into account (superficial lesions were not clearly identifiable). Foot-pad dermatitis is considered a useful indicator of animal welfare and our findings suggest that visual, macroscopic scoring could be a practicable assessment tool.*

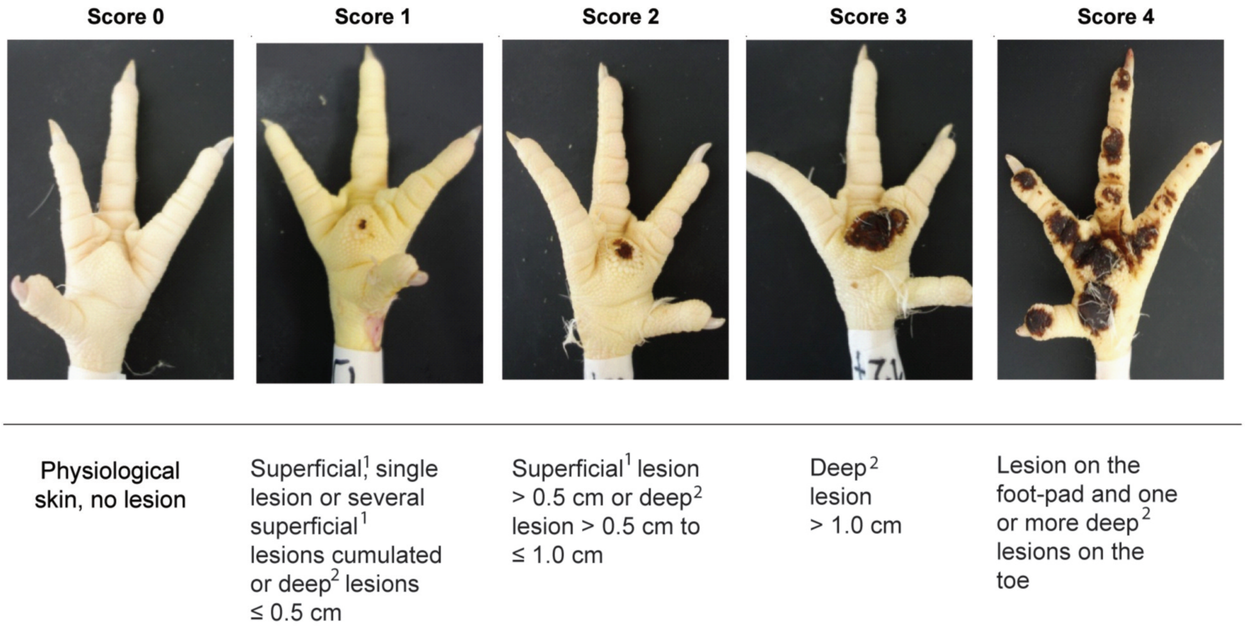
Keywords: *animal welfare, animal welfare indicator, broiler, foot-pad dermatitis, histological validation, poultry*

Introduction

Foot-pad dermatitis (FPD) is a contact dermatitis on the plantar surface of the foot pad in broilers which can cause acute inflammatory and necrotic lesions (Greene *et al* 1985; Shepherd & Fairchild 2010). 'Hock burn' and 'breast burn' are similar forms of contact dermatitis, showing the same clinical signs in poultry as FPD but appearing on other parts of the body (Haslam *et al* 2007). The severity of FPD can depend on genetics (Ask 2010), stocking density (Bruce *et al* 1990; Spindler & Hartung 2011) and nutrition (Martland 1985), but the primary cause is wet litter (Martland 1985; Meluzzi *et al* 2008; Weber Wyneken *et al* 2015). Since FPD is a common occurrence in the conventional poultry industry (Saraiva *et al* 2016) and, almost certainly, painful (Algers & Berg 2001), it is a useful indicator of animal health and welfare (Allain *et al* 2009).

First described in the 1980s (Greene *et al* 1985; Martland 1985), this has been the subject of several studies throughout recent decades and several authors have had various scoring systems published that have sought to categorise FPD (Greene *et al* 1985; Martland 1985; Ekstrand *et al* 1997, 1998; Bilgili *et al* 2006; Welfare Quality® 2009; McKeegan 2010). As a result there is no fixed and uniform system currently in use (Heitmann *et al* 2018). The various systems differ in terms of the numbers of categories but, in most, a three- (Ekstrand *et al* 1998; Bilgili *et al* 2006; Welfare Quality® 2009) or a four-point scale (Martland 1985; Martrenchar *et al* 2002) is used to describe the macroscopic findings. Furthermore, Martland (1985), Greene *et al* (1985) and McKeegan (2010) considered the FPD lesions histopathologically, and Michel *et al* (2012) described FPD via a five-point scaled histological score, drawing a link to macroscopic results. The five-point

Figure 1



Macroscopic scores for the visual assessment of foot-pad dermatitis based on a modified Welfare Quality® (2009) assessment scheme. 1 = Lesions with discolouration, hyperkeratosis, no visual disruption of epidermal layers; 2 = Lesions with loss of scales, visual disruption of epidermal layers

scaled histological scoring system for hock burn and FPD published by McKeegan (2010) contains the following lesion scores: (A) unaffected skin; (B) early lesion with mild hyperkeratosis and a mild disruption of keratin on the scale; (C) intermediate lesion with acanthosis, increased vascularity and hyperkeratosis; (D) advanced lesion with ulcer, serocellular crusting and exudation of heterophils; and (E) advanced lesion with extensive ulcer, serocellular crusting and granulation tissue.

Excluding normal feet and foot pads with scars or healing lesions, three types of lesions were described histologically by Michel *et al* (2012): type I, mild epidermal hyperplasia and/or hyperkeratosis; type II, marked hyperkeratosis and hyperplasia of the epidermis with marked dermal inflammation; and type III, ulceration with loss of epidermis. In Germany, there is currently an officially recommended system for visual scoring of FPD macroscopically: the German Order on the Protection of Animals and the Keeping of Production Animals (2006). Additionally, a modified version of this, implementing a four- instead of a three-point scale, was launched by the Implementary Rules of the Lower Saxonian Ministry of Nutrition (2015) and is in use in Lower Saxony, Germany. FPD evaluation in German slaughterhouses is randomly monitored by veterinarians who visually score 100 feet per flock. In some plants, cameras are used to scan for FPD (Louton *et al* 2018). From 2013 to 2015, 32% of EU broiler production, in approximately one-third of member states, were scoring FPD (European Commission Directorate-General for Health and Food Safety 2016).

The aim of this study was to determine whether a link exists between histological (ie microscopic) and visual (ie macroscopic) FPD scores. Furthermore, the relationship between these scores and the histologically measured depth of inflammation was assessed.

Materials and methods

Initially, a scoring system to categorise FPD was chosen and inter-observer reliability was assessed. Then, broiler feet were collected and photographed, before a veterinarian measured the size of the lesions and scored feet. Histological slides were then produced allowing inflammation and foot lesions to be scored and measured.

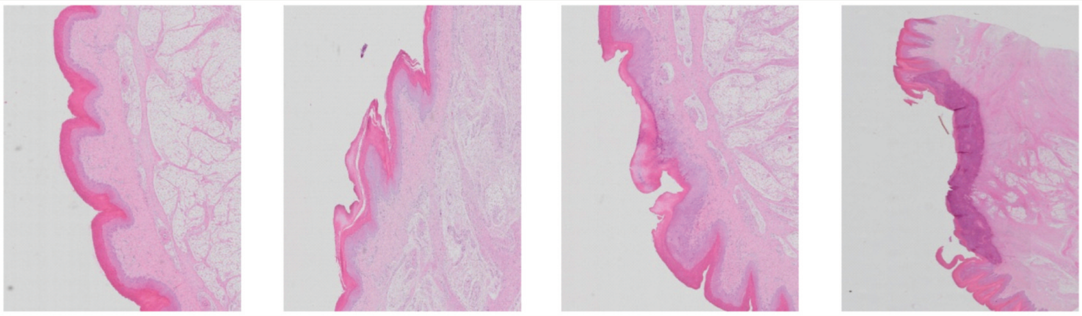
All examined feet were from Ross 308 broilers aged 39–42 fattening days and taken from the same broiler slaughterhouse: Donautal Geflügelspezialitäten Zweigniederlassung der Lohmann & Co KG in Bogen, Germany.

Inter-observer reliability

The aim was to choose an easy-to-use assessment scheme with a high inter-observer reliability for categorising of FPD scores. After an evaluation of pre-existing schemes, it was decided to use a modified version (Figure 1) of the Welfare Quality® Assessment Protocol for Poultry (2009). To validate the inter-observer reliability, 250 broiler feet were collected from the slaughterhouse between July and November 2017. Feet were stored at −18°C prior to use. The intention was to distribute scores as equally as possible and, after the samples were defrosted, five observers scored them independently and in a random order. Although assessing feet visually, observers were invited to use a

https://doi.org/10.7120/09627286.29.2.185 Published online by Cambridge University Press

Figure 2



Score 0	Score 1	Score 2	Score 3
Physiological skin, no lesion	Mild to moderate hyperplasia and/or hyperkeratosis of the epidermis; superficial dermal congestion and oedema, variable and mild inflammatory infiltration	Marked hyperplasia and hyperkeratosis of the epidermis, abundant heterophilic epidermal exocytosis with pustule formation; congestion in the superficial dermis with inflammatory infiltration	Ulceration (full thickness necrosis of the epidermis, replaced by necrotic and suppurative material)

Microscopic scores for foot-pad dermatitis, modified according to Michel *et al* (2012).

Precise PS 7215 digital measuring stick (Burg-Wächter KG, Wetter, Germany). Inter-observer reliability was calculated using the prevalence- and bias-adjusted kappa (PABAK/kappa$_{nor}$) (Byrt *et al* 1993). According to Gunnarsson *et al* (2000), the following modified PABAK should be used when observers can assign more than two categories (k = number of categories; p_0 = relation between observed agreements):

PABAK = $(kp_0 - 1)/(k - 1)$

In addition, three veterinary pathologists from the Department of Pathology, Bavarian Animal Health Service, Poing, Germany, who scored all of the feet histologically, performed an inter-observer comparison with 20 randomly picked FPD slides and scored these individually. The PABAK values and the inter-observer reliabilities were calculated as per the first observer comparison.

Data collection, applied macroscopic and applied microscopic scores

Next, 200 feet (40 per macroscopic score) were obtained from the slaughterhouse and macroscopically scored according to Figure 1. Each foot was photographed (Sony Cyber-shot DSC-RX100 camera, Sony Europe Limited, Surrey, UK) and total foot pad and FPD sizes were measured directly on each foot, using the length between the most distant points. The feet were then taken to the Department of Pathology, Bavarian Animal Health Service, Poing, Germany and cut at the most critical point as determined visually. Samples were fixed in 10% buffered formalin for 24 h and then embedded in paraffin wax. Sections of 4-µm thickness were mounted onto glass slides, stained with haematoxylin and eosin, examined histologically and scored by veterinary pathologists using a modified version of Michel *et al*'s scheme (2012) (Figure 2). Additionally, each foot was examined histologically for potential pre-existing tendon inflammation.

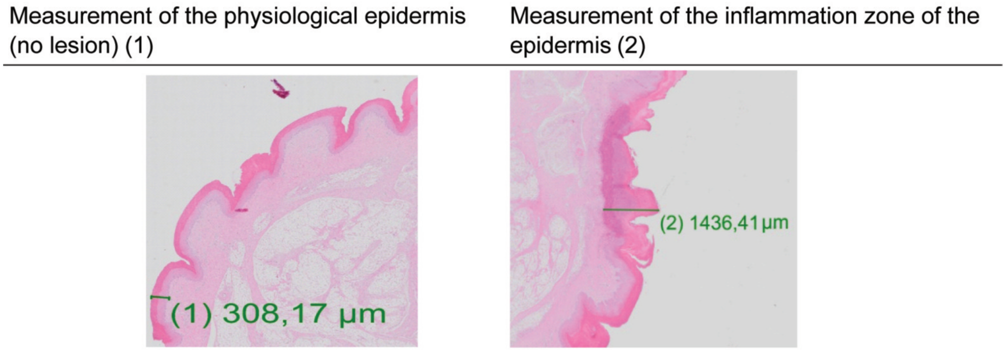
Histological samples were digitalised by using an Olympus BX51 microscope (Olympus, Tokyo, Japan). Furthermore, physiological skin depth from stratum corneum to stratum basale and the depth of the inflammation zone from stratum corneum to the demarcation line (if present), were measured using the Olympus VS-ASW (Olympus, Tokyo, Japan) (Figure 3).

Downscaled microscopic score

In addition to the classification according to our modified four-point microscopic score (based on Michel *et al* 2012; Figure 2), a downscaled, three-point score was used simultaneously. For this, microscopic scores 1 and 2 were merged. Score 0 is identical in both schemes and represents feet with 'no lesion'. Score 1 in the downscaled microscopic score denotes 'mild lesion' and includes our modified microscopic scores 1 and 2. Downscaled microscopic score 2 represents ulcers and is identical to our modified microscopic score 3.

doi: 10.7120/09627286.29.2.185

Figure 3



Examples of the histological measurements for the depth of the inflammation zone of the epidermis.

Table 1 PABAK results of the analysis of inter-observer reliability for the macroscopic scoring of foot-pad dermatitis (five observers; n = 250 feet).

Compared observer	1/2	2/3	1/3	2/4	3/4	1/4	2/5	3/5	4/5	1/5	Average
Match	225	226	223	228	228	221	227	227	227	226	
No match	25	24	27	22	22	29	23	23	23	24	
%	0.90	0.90	0.89	0.91	0.91	0.88	0.91	0.91	0.91	0.90	0.90
PABAK	0.88	0.88	0.87	0.89	0.89	0.86	0.89	0.89	0.89	0.88	0.88

Macroscopic score regarding depth

After data had been collected, two veterinarians, who had also taken part in the first inter-observer comparison, visually scored the 200 pictures of the study feet, assessing only the depth of the inflammation. Scoring took place without any prior knowledge of the original macroscopic score as per Figure 1. The macroscopic scoring system addressing the depth of inflammation was as follows: 0 = no lesion; 1 = superficial lesions with discolouration, hyperkeratosis, no visual disruption of epidermal layers; and 2 = deep lesions with loss of scales, visual disruption of all epidermal layers.

Statistical analysis

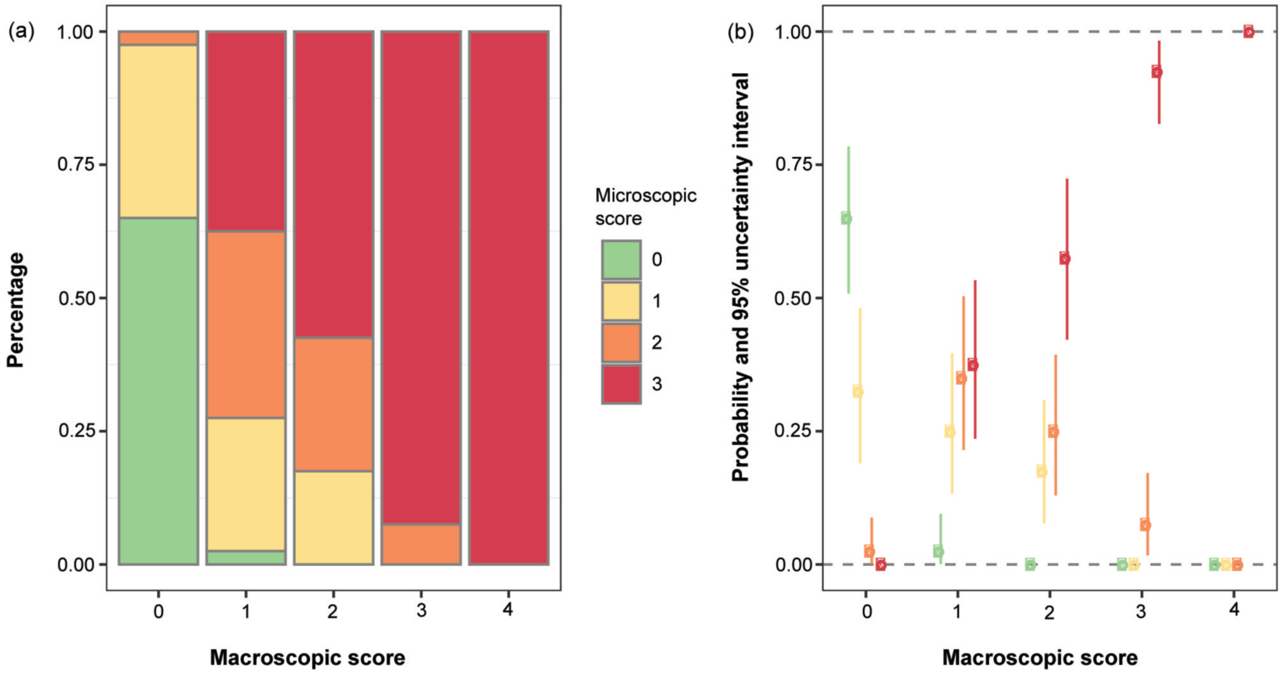
To assess level of agreement between microscopic and macroscopic scores, the conditional probabilities of all microscopic categories given the macroscopic categories were estimated via multinomial regression models. To examine the relationship between macroscopic score and depth of inflammation, a Hurdle-Gamma regression model was used. Here, the Hurdle partly models the conditional probability for an inflammation depth > 0 µm given the macroscopic categories and the Gamma partly investigates the effect of the macroscopic score on the expected depth of inflammation > 0 µm. The Gamma distribution was chosen to account for the strictly positive and therefore skewed distribution of inflammation depth.

Prior to data collection, the required sample size was estimated. To this end, data for microscopic and macroscopic scores were simulated by assuming conditional probabilities of all microscopic categories given the macroscopic categories. Data were generated according to three different sets of conditional probabilities (optimistic, neutral and pessimistic) affecting the separability of score categories. Furthermore, equal sample size along the macroscopic categories was assumed. The generated datasets were analysed using the same multinomial regression model as above. Given the fitted model, it was checked if all hypothetically assumed conditional probabilities were within the estimated credible intervals. This step was repeated 10,000 times for each set of conditional probabilities and for sample sizes ranging from 20 to 100. The resulting proportion of simulations in which all conditional probabilities were estimated correctly was interpreted as statistical power.

All model parameters were estimated in a fully Bayesian way using the probabilistic programming language Stan (Carpenter *et al* 2017) and the wrapper package brms (Bürkner 2017) for the statistical programming language R (R Core Team 2018). For all estimated model parameters, point estimates as well as 95% credible/uncertainty intervals are reported. For the evaluation of cut-off values and the classification of attributes, the following performance measures were used: accuracy, sensitivity, specificity, positive and negative predictive value.

https://doi.org/10.7120/09627286.29.2.185 Published online by Cambridge University Press

Figure 4



Showing (a) descriptive distribution of microscopic scores (0–3) for given macroscopic scores (0–4) of foot-pad dermatitis in broilers (Ross 308; 200 feet) and (b) probabilities and 95% uncertainty intervals of microscopic scores (0–3) for given macroscopic scores (0–4) of foot-pad dermatitis in broilers (Ross 308; 200 feet).

Results

Inter-observer reliability

As detailed in Table 1, the relations between matching of observed agreements varied from PABAK values of 0.86 up to 0.89, with an average of 0.88. The PABAK values of the second inter-observer comparison based on the histological findings were 0.80, 0.93 and 0.87, with an average of 0.87.

Relation between macroscopic and microscopic score

Considering the macroscopic scores, lower or higher macroscopic scores corresponded, respectively, to lower or higher microscopic scores, although not on a one-to-one basis. Figure 4(a) shows the descriptive distribution of microscopic scores at given macroscopic FPD scores. The macroscopic scores contained mixed microscopic scores, with the exception of macroscopic score 4, which contained 100% microscopically ulcer (microscopic score 3).

Nearly all feet with microscopically no lesion (score 0) were visually diagnosed with no visual lesion (96.3%, macroscopic score 0), apart from one foot that was diagnosed with macroscopic score 1.

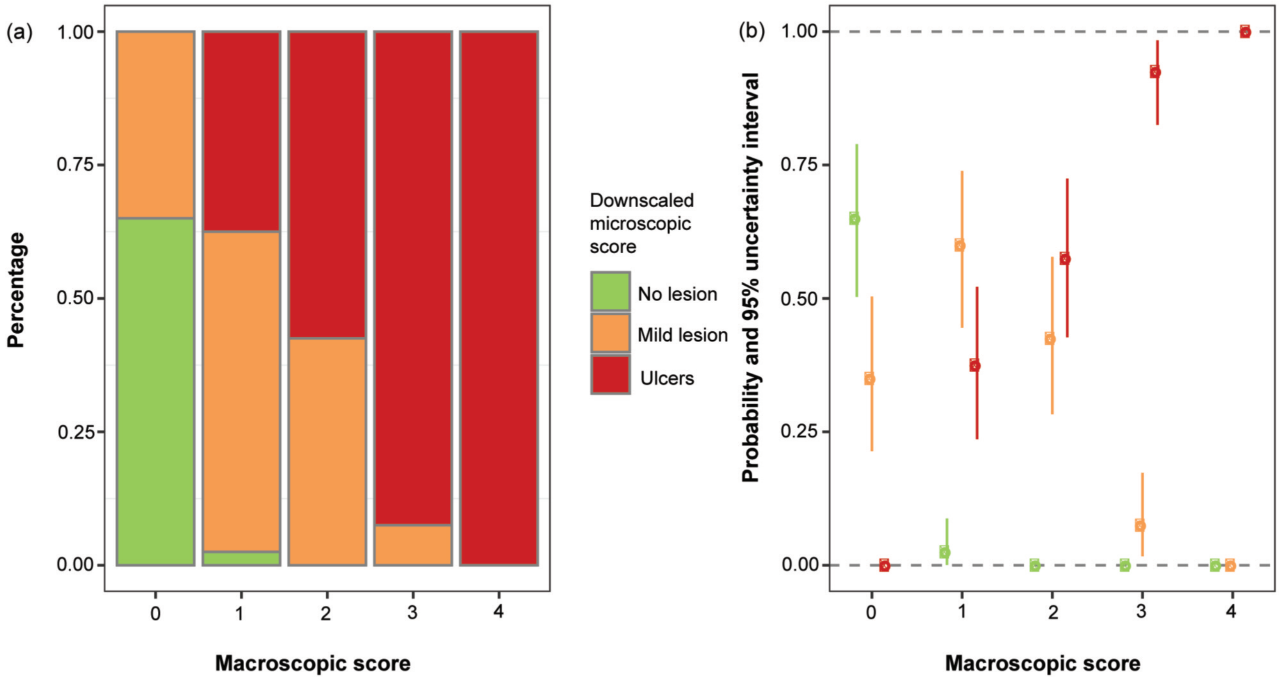
Figure 4(b) presents the results of the conditional probabilities of respective microscopic scores given macroscopic scores as well as the corresponding 95% credible/uncertainty intervals. For macroscopic score 0 (no visual lesion), microscopic score 0 (0.650, no lesion) had the highest probability. For macroscopic score 1, the probabilities for microscopic scores 1 (0.250), 2 (0.350) and 3 (0.375) were very similar and microscopically no lesions were found with low probability (0.025). For macroscopic score 2, no probability for microscopic score 0, moderate probabilities for microscopic scores 1 (0.175) and 2 (0.250) and high probability for microscopic score 3 (0.575) were found. Macroscopic score 3 showed no probability for microscopic scores 0 and 1 (0.000), low probability for microscopic score 2 (0.075) and very high probability for microscopic score 3 (0.925, ulcer). Macroscopic score 4 only showed probability for microscopic score 3 (1.000, ulcer).

Relation between macroscopic score and downscaled microscopic score

Figure 5(a) presents the relationship between macroscopic scores and downscaled microscopic scores (no lesion, mild lesion and ulcer). Macroscopic score 0 (macroscopically no lesion) contained 65.0% feet with microscopic no lesion and 35.0% feet with mild lesion. Macroscopic score 1 contained 2.5% feet with no lesions, 60% mild lesions and 37.5% ulcer. Macroscopic score 3 contained 7.5% mild lesion and 92.5% ulcer. Macroscopic score 4 contained only ulcer (100.0%).

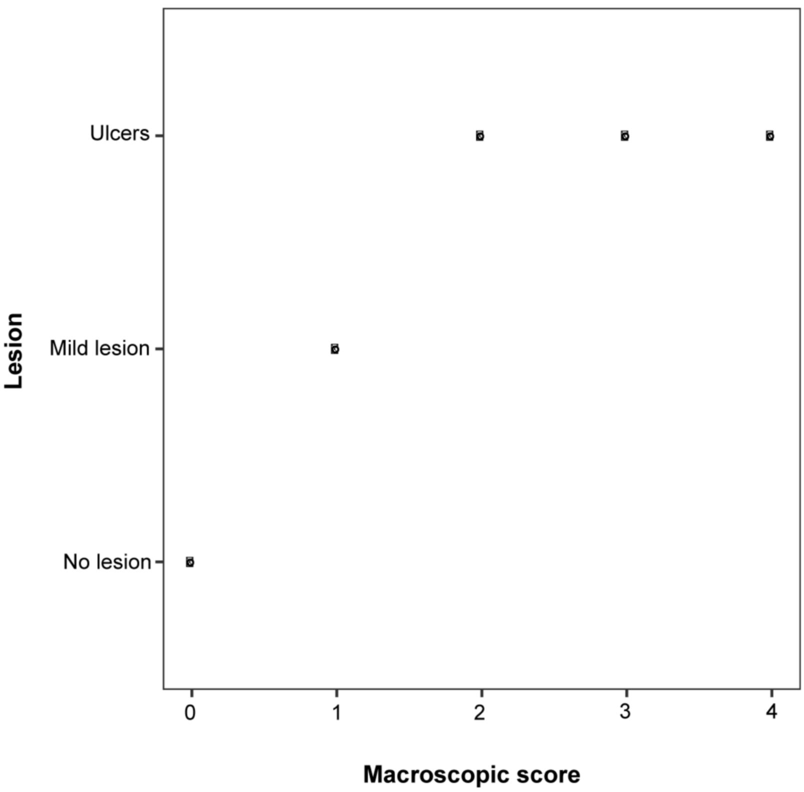
Nearly all feet with mild lesions were visually diagnosed with macroscopic score 1 (96.3%). The maximum peak for mild lesions was reached at macroscopic score 1. Along the increasing scale of macroscopic scores, a moderate decrease was seen in the prevalence of mild lesions (downscaled microscopic score 1) and a moderate increase in the prevalence of ulcer (downscaled microscopic score 2), although there was no one-to-one assignment to either.

Figure 5



Showing (a) descriptive distribution of downscaled microscopic scores (no lesion, mild lesion, ulcers) for given macroscopic scores (0–4) of foot-pad dermatitis in broilers (Ross 308; 200 feet) and (b) probabilities and 95% uncertainty intervals of downscaled microscopic scores (no lesion, mild lesion, ulcers) for given macroscopic scores (0–4) of foot-pad dermatitis in broilers (Ross 308; 200 feet).

Figure 6



Highest probability of downscaled microscopic scores (no lesion, mild lesion, ulcers) at the respective macroscopic scores (0–4) of foot-pad dermatitis in broilers (Ross 308; 200 feet).

https://doi.org/10.7120/09627286.29.2.185 Published online by Cambridge University Press

Table 2 Performance measures of the macroscopic categories for the prediction of microscopic scores of foot-pad dermatitis in broilers (Ross 308; 200 feet).

(a)	Prediction			Performance			
	Macroscopic = 0	Macroscopic = 1	Macroscopic ≥ 2	Accuracy = 0.75			
Microscopic	No lesion	Mild lesion	Lesion with ulcer	Sens	Spec	PPV	NPV
No lesion	26	1	0	0.96	0.92	0.65	0.99
Mild lesion	14	24	20	0.41	0.89	0.60	0.79
Lesion with ulcer	0	15	100	0.87	0.76	0.83	0.81
(b)	**Prediction**			**Performance**			
	Macroscopic = 0	**1 ≤ Macroscopic ≤ 2**	**Macroscopic > 2**	**Accuracy = 0.72**			
Microscopic	**No lesion**	**Mild lesion**	**Lesion with ulcer**	**Sens**	**Spec**	**PPV**	**NPV**
No lesion	26	1	0	0.96	0.92	0.65	0.99
Mild lesion	14	41	3	0.71	0.73	0.51	0.86
Lesion with ulcer	0	38	77	0.67	0.96	0.96	0.68

Sens = Sensitivity: the probability that an animal belonging to category c is identified as category c;
Spec = Specificity: the probability that an animal not belonging to category c is identified as non-category c;
PPV = Positive predictive value: the probability that an animal is identified as category c and belongs to it;
NPV = Negative predictive value: the probability that an animal is not identified as category c and does not belong to category c;
Accuracy = Probability for the correct classification of the animals.

Figure 5(b) shows the conditional probabilities with corresponding uncertainty intervals of downscaled microscopic scores given the macroscopic scores. A probability for downscaled microscopic score 0 (no lesion) could only be found in feet with macroscopic scores 0 (0.650) and 1 (0.025). High probabilities of mild lesions were observed in feet with macroscopic score 1 (0.600) and 2 (0.425) and for ulcer in feet with macroscopic scores 2 (0.575), 3 (0.925) and 4 (1.000).

Cut-off data and resulting performance values

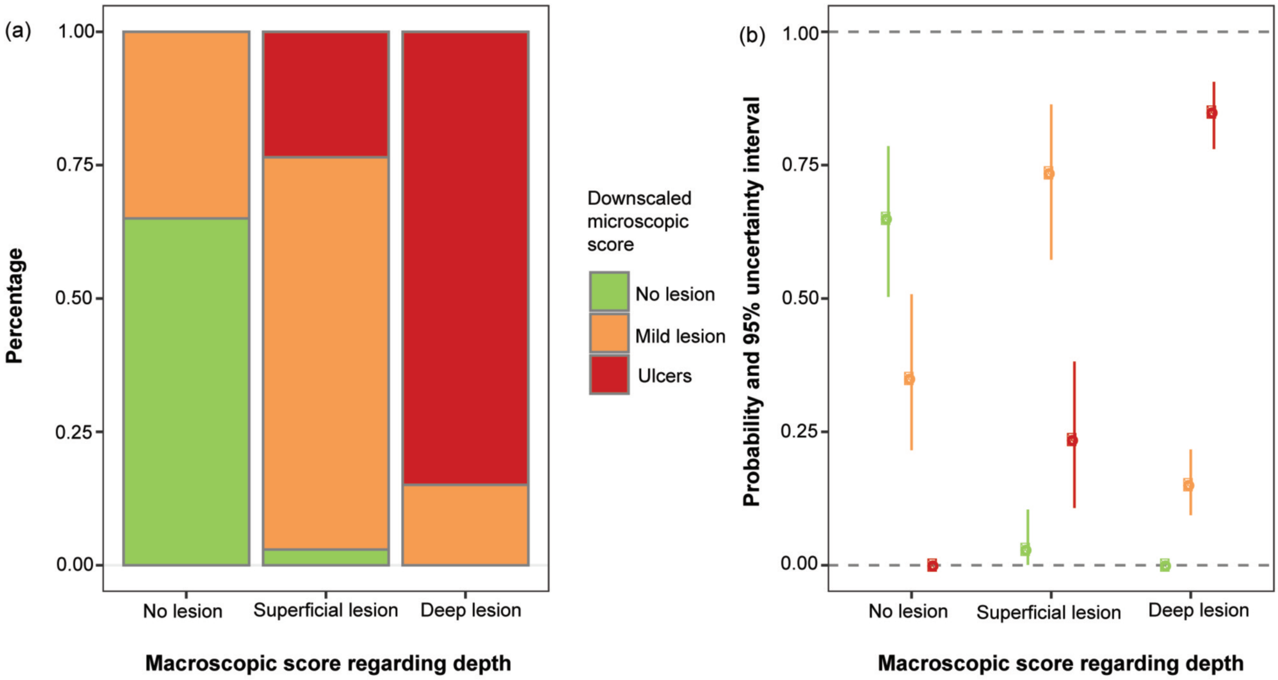
Figures 4(b) and 5(b) clearly indicate which microscopic or downscaled microscopic score had the highest probability for each macroscopic score in the given macroscopic score scheme. In Figure 6, the downscaled microscopic score with the highest probability is shown for the given macroscopic scores. At macroscopic score 0, no lesion (downscaled microscopic score 0) showed the highest probability. At macroscopic score 1, mild lesions (downscaled microscopic score 1) had the highest probability of diagnostic. For all other macroscopic scores (2, 3 and 4), ulcer (downscaled microscopic score 2) showed the highest probability. Thus, cut-off values can be calculated. The resulting performance values given a cut-off value for 'no lesion' at macroscopic score 0, 'mild lesion' at macroscopic score 1 and 'ulcer' at macroscopic score 2 can be seen in Table 2(a). In contrast, in Table 2(b), macroscopic scores 1 and 2 were categorised as microscopic 'mild lesion' and macroscopic score 3 and above as microscopic 'ulcer'. For the first calculated cut-off values (Table 2[a]), 26 out of 27 feet with no lesions were assigned to macroscopic score 0 and thus could be diagnosed with a sensitivity of 0.96 and specificity of 0.92. Feet with mild lesions could be diagnosed with a sensitivity of 0.41 and specificity of 0.89. One hundred out of the 115 feet with ulcer were diagnosed with macroscopic score ≥ 2, resulting in a sensitivity of 0.87 and specificity 0.76. Given the cut-off value for ulcer at macroscopic score 2, the accuracy of all performance values was 0.75.

On the other hand, when macroscopic scores 1 and 2 were categorised as mild lesions resulting in a cut-off value for ulcer at macroscopic score 3 (Table 2[b]), all performance values were calculated with an accuracy of 0.72. Similarly to the results in Table 2(a), 26 out of 27 feet with no lesion were assigned to macroscopic score 0 and thus could be diagnosed with a sensitivity of 0.96 and specificity of 0.92. Feet with mild lesions could be predicted with a sensitivity of 0.71 and specificity of 0.73. Feet with ulcer could be predicted with a sensitivity of 0.67 and specificity of 0.96. In addition, the positive and negative predictive values were calculated for the diagnosis of the downscaled microscopic score (Table 2[a] and [b]).

Relation between macroscopic score regarding depth and downscaled microscopic score

Figure 7(a) shows the descriptive distribution of macroscopic scores considering the depth of inflammation in relation to the downscaled microscopic scores. Feet with no visible lesion (macroscopic score regarding depth 0) included mainly feet with no histological lesion (downscaled microscopic score 0), and ulcer were not detected. Mostly mild lesions were observed in feet with visually

Figure 7



Showing (a) descriptive relation between downscaled microscopic scores (no lesion, mild lesion, ulcers) and macroscopic scores considering the depth of inflammation (no lesion, superficial lesion, deep lesion) of foot-pad dermatitis in broilers (Ross 308; 200 feet) and (b) probabilities and 95% uncertainty intervals of downscaled microscopic scores (no lesion, mild lesion, ulcers) for given macroscopic scores regarding depth (no lesion, superficial lesion, deep lesion) of foot-pad dermatitis in broilers (Ross 308; 200 feet).

superficial lesions (macroscopic score regarding depth 1). Visually deep lesions (macroscopic score regarding depth 2) included mainly ulcer.

Figure 7(b) shows the different probabilities with 95% uncertainty intervals for the macroscopic scores regarding depth and the downscaled microscopic scores (no lesion, mild lesion, ulcers). Feet with no visual lesion had the highest probability for histological diagnosis with no lesion (0.650). Visual superficial lesions (macroscopic score regarding depth 1) had the highest probability for diagnosis with mild lesions (0.735) and the lowest for no lesions (0.003). Visually deep lesions (macroscopic score regarding depth 2) had very high probability for diagnosis with ulcer (0.850), low probability for mild lesions (0.150) and none for no lesions (0.000).

Relation between macroscopic score and histological depth of inflammation

The histologically measured depth of inflammation in lesions of the foot pad is presented for the different macroscopic scores (0, 1, 2, 3 and 4) in a boxplot (Figure 8[a]). An increase in the depth of inflammation with increasing macroscopic score up to macroscopic score 3 was found, with mean values increasing from 937.1 (macroscopic score 1) to 1,087.4 (macroscopic score 2) to 1,443.0 µm (macroscopic score 3). The mean value for macroscopic score 0 was 8.0 µm, and 97.5% of all feet assigned to macroscopic score 0 were measured with 0.0 µm (with no histological lesion). Feet assigned to macroscopic score 4 showed a mean inflammation depth of 1,422.2 µm, slightly less than those assigned to macroscopic score 3. Macroscopic scores 1 and 2 included 27.5 and 12.5% feet, respectively, with 0.0-µm depth of inflammation. A strong decrease in feet was noted with 0.0-µm depth of inflammation with increasing macroscopic score up to macroscopic score 2. The boxplot in Figure 8(a) shows macroscopic scores 1 to 4 to have similar 0.50 quantiles and to vary in the space between 0.25 and 0.75 quantiles.

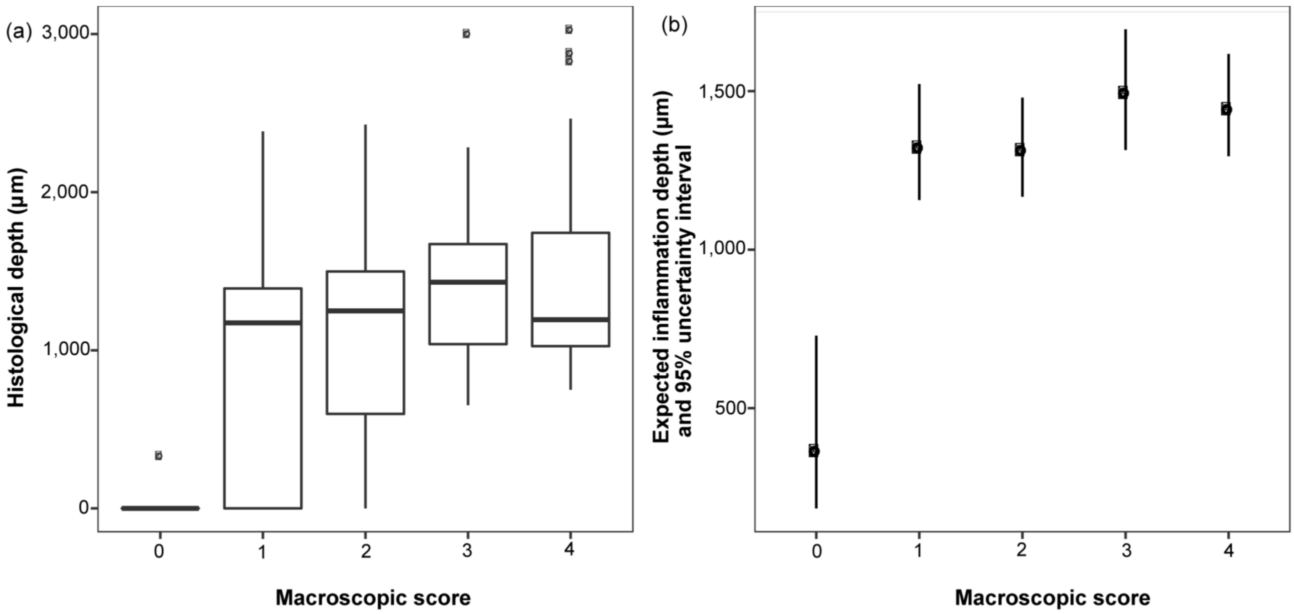
The relation between macroscopic scores and expected inflammation depth was further analysed by a regression model (Hurdle-Gamma). The estimated expected inflammation depth is shown in Figure 8(b). This analysis revealed significant differences between macroscopic score 0 and the remaining macroscopic scores (1 to 4). In all cases, macroscopic score 0 showed the significantly smallest expected inflammation depth. Differences between higher macroscopic scores could not be confirmed.

Tendinitis

In feet with microscopic scores 0, 1 and 2, no tendinitis was observed. The two feet (out of 200 examined feet) with tendinitis were scored as microscopic score 3 (ulcer). As both feet were categorised as macroscopic score 4, there was no inflammation of the tendons detected in macroscopic scores 0 to 3.

https://doi.org/10.7120/09627286.29.2.185 Published online by Cambridge University Press

Figure 8



Showing (a) boxplot of relation between macroscopic scores (0–4) and the histologically measured depth (µm) of inflammation of foot-pad dermatitis in broilers (Ross 308; 200 feet) and (b) expected inflammation depth (µm) and 95% uncertainty intervals for respective macroscopic scores (0–4) of foot-pad dermatitis in broilers (Ross 308; 200 feet).

Discussion

Inter-observer reliability

The PABAK values of the inter-observer comparison for the rating of both the macroscopic and the histological findings ranged from 0.81 to 1.00. According to Landis and Koch (1977), this range represents an almost perfect reliability score.

Macroscopic score mirroring histological findings with limitations

Ulcerations are defined as lesions with perforated and destroyed epidermis including stratum basale, whereas mild lesions show hyperkeratosis, hyperplasia or mild inflammation with intact stratum basale (Michel *et al* 2012). In this study, we used a visual, macroscopic scoring system for FPD with and without considering the depth of inflammation in the observed lesion and compared it with a histological, microscopic scoring. Our histological findings identified the macroscopic scores regarding depth as most suitable predictors of the histopathological presence of mild lesions or ulcers, taking various limitations into account. In general, the results illustrated a clear relationship between macroscopic scores and histopathological findings. For the macroscopic examinations, feet with no histologically detectable lesions were mainly assigned to feet with no visually detected lesions; similarly, feet with histologically identified ulcers were related to feet with visually scored bigger and deeper lesions.

In line with Michel *et al* (2012), our analyses revealed a reliable link between macroscopic and microscopic scoring with a few exceptions. Mayne *et al* (2006) showed similar results in turkeys and observed that feet with externally normal foot pads revealed histopathological evidence of lesions. Our results confirm that histologically detectable mild lesions with hyperkeratosis and/or erosions can be distributed widely throughout feet with visually no detected lesions, mild/superficial lesions and big/deep lesions. Thus, microscopic scores 1 and 2 (mild lesions) are not easy to distinguish visually by use of a macroscopic score from feet with no lesions or ulcers. This limitation may explain the poor performance values for the diagnosis of mild lesions when only a macroscopic score was received (ie when macroscopic score 2 was the cut-off value to assign ulcers). However, the diagnoses of feet with no lesions and feet with ulcers had high performance values in this scenario, and these feet could be assigned macroscopic scores with high probabilities.

In contrast, when macroscopic scores 1 and 2 were categorised as mild lesions (ie when macroscopic score 3 was the cut-off value to assign ulcers), we found a much higher sensitivity (0.71 instead of 0.41) and a lower specificity (0.73 instead of 0.89) for the diagnosis of mild lesions. Thus, mild lesions were more likely to be diagnosed correctly. However, lesions with ulcers (macroscopic scores 3 and 4) could then be diagnosed with a lower sensitivity (0.67 instead of 0.87) and a higher specificity (0.96 instead of 0.76) compared with the first presented cut-off value.

To summarise, if the focus is on detecting ulcers, high sensitivity (albeit low specificity) would be achieved by using the first cut-off value. The second cut-off value would allow a better detection of mild lesions. The accuracy for the first cut-off value (0.75) was higher than for the second (0.72).

The results of this study confirm the correlation between macroscopic and histological findings reported by

McKeegan (2010) and Michel *et al* (2012). In the study of McKeegan (2010), feet with visually scored lesions mainly had histologically detectable inflammatory lesions although a number of exceptions were noted. Similarly, we found that certain feet with no visual evidence of lesions showed mild lesions in the histological examination.

Importance of identifying ulcers

In the scoring system of the German Order on the Protection of Animals and the Keeping of Production Animals (2006), the highest score (2b) is defined as the presence of an ulcer. In addition, it is generally accepted that FPD, particularly in the form of severe ulcers, is likely to cause pain (Algers & Berg 2001). However, to date, no study has differentiated or categorised grades of FPD or inflammation depths in terms of different pain levels or been able to clearly define which grade or inflammation depth is relevant as regards animal welfare. Michel *et al* (2012) supported the hypothesis that the level of pain increases relative to the size of the lesion and according to the different levels of severity (depth of inflammation) seen between microscopic scores 2 and 3. The authors drew on the work of Arnould and Colin (2008) to formulate their conclusion and the enhanced reaction to manipulation with increasing severity of FPD. Hocking and Wu (2013) also adhered to the notion that FPD is likely to cause pain in turkeys. Moreover, McKeegan (2010) found chickens with FPD to move more slowly than those with no lesions and walk better with analgesia than without. Martland (1985) also drew a possible link between decreased appetite associated with pain and reduced growth rate. Furthermore, Weber Wyneken *et al* (2015) found a significant interaction between FPD and analgesia regarding stride length in turkeys and differences between birds with low- and high-grade FPD. However, lameness in broilers may not be related to FPD and can occur for other reasons, such as chondronecrosis and osteomyelitis (Wideman 2016).

Until scientific research confirms a difference in pain level according to different grades of FPD, an existing ulcer with inflammation should be considered painful (Algers & Berg 2001). Following this assumption, a macroscopic scoring system for FPD able to differentiate feet with no histological lesion from feet with ulcerations should be used to record FPD as an animal welfare indicator. The probability for FPD with an ulcer is highest at macroscopic scores 2 and above in our scoring system. Thus, a lesion with a diameter greater than 0.5 cm is likely to be a deep lesion with an ulcer and considered painful.

Overall, to diagnose an ulcer, we can recommend a simplified three-point macroscopic scoring system for FPD to differentiate between feet with no lesion, feet with mild lesions and feet with ulcers. Several, similar, three-point scores have been published, such as the summarised score of the Welfare Quality® Assessment Protocol for Poultry (2009) or that of Heitmann *et al* (2018). Considering, in particular, a camera-based and standardised detection of FPD, a three-point scale that can clearly distinguish an ulcer from other lesions would be applicable. Based on our results, a good suggestion for a useful score would be:

- Score 0 = no lesion;
- Score 1 = mild lesions (superficial lesions with discolouration, hyperkeratosis and no visual disruption of the epidermal layers, size ≤ 0.5 cm in diameter);
- Score 2 = deep lesions with loss of scales, visual disruption of all epidermal layers, ulcers and/or crusting (size > 0.5 cm in diameter).

A similar scoring system has been published by Ekstrand *et al* (1998), however the authors did not define specific lesion sizes. In the commonly used scoring systems at slaughter, macroscopic scores are used to categorise FPD (Louton *et al* 2018), and according to our results, a routine histological examination is not necessary to score FPD effectively. Our study shows that it is possible to extrapolate from the macroscopic scoring of FPD to the type of lesion, with certain limitations.

Depth of inflammation and tendinitis

Although no significant statistical differences were found between the different macroscopic scores according to the histologically determined depth of inflammation, we observed an increasing inflammation depth with increasing macroscopic score. Mirroring the results of Heitmann *et al* (2018), a positive correlation was found between inflammation depth and size of the lesion in ulcers. However, since Heitmann *et al* (2018) measured the total thickness of skin layers (while we measured the depth of the inflammation zone), a one-to-one comparison of the results is not possible. Heitmann *et al* (2018) described large ulcers (90 mm^2) with an increase in thickness of skin layers to 1,535.5 μm, a value similar to the mean depth of the inflammation zone in macroscopic scores 3 (1,443.0 μm) and 4 (1,422.2 μm) measured in our study.

Another way of scoring FPD could be via the development of a practical and non-invasive technique for detecting the depth of inflammation at the slaughterhouse, eg by using infra-red cameras (Wilcox *et al* 2009). Such an approach might increase the likelihood that mild lesions would be diagnosed with greater accuracy.

Affected tendons were found only in individual cases and no significant relation was discovered between FPD and detectable inflammation of tendons.

Animal welfare implications

The validated scoring systems discussed here can contribute to the improvement of animal welfare within the poultry industry. Taking into account our results, we recommend a three-point, visual scoring system for the detection of ulcers in FPD. Moving forward, a macroscopic, comprehensive and standardised detection of FPD could be especially beneficial in improving the health and welfare of broiler chickens. Future research should evaluate if the results of this study can also be applied to an automatic scoring system with camera-based detection of FPD. Furthermore, more research is needed to consider the perception of pain at different inflammation depths or grades of FPD and additionally future research should focus on reducing the prevalence of FPD. Nevertheless, this histologically validated, veterinarian-endorsed confirmation of macroscopic scoring and inflammation depth of FPD lend validity to these macroscopic scores operating as indicators of animal welfare.

https://doi.org/10.7120/09627286.29.2.185 Published online by Cambridge University Press

Conclusion

In general, our study elaborates significant links between the examined macroscopic and microscopic scoring systems and demonstrates the extent of their comparability. However, limitations, notably the diagnosis of mild lesions such as hyperkeratosis, erosions and other superficial lesions need to be considered. A routine examination based upon visual scoring can be recommended for the diagnosis of ulcers; it accurately mirrors histopathological findings, especially when feet with no lesions and feet with ulcers are the main focus. Although no statistically significant differences in inflammation depth between the visually present lesions were found, an increase in inflammation depth with increasing severity of FPD was detected. A link between tendinitis and FPD was not found.

Acknowledgements

This project was supported by funds from the Federal Ministry of Food and Agriculture (BMEL) based on a decision from the Parliament of the Federal Republic of Germany via the Federal Office for Agriculture and Food (BLE) under the innovation support programme (grant number: FKZ 2817903715). The funding source played no role in the study design; in the collection, analysis and interpretation of data; in the writing of the report; or in the decision to submit the paper for publication. We thank Dr Sven Reese for providing the Olympus BX51 microscope and great scientific support, Dr Verena Lietze for editing of the scientific language as well as Stefan and Yeliz Wiesbeck, Astrid Nagel and Birgit Isele-Rüegg for excellent technical assistance. There is no conflict of interest regarding any financial, personal or other relationships with other people or organisations that could inappropriately influence or be perceived to influence this study.

References

Algers B and Berg C 2001 Monitoring animal welfare on commercial broiler farms in Sweden. *Acta Agriculturae Scandinavica, Section A, Animal Science 51*: 88-92. https://doi.org/10.1080/090647001316923135

Allain V, Mirabito L, Arnould C, Colas M, Le Bouquin S, Lupo C and Michel V 2009 Skin lesions in broiler chickens measured at the slaughterhouse: Relationships between lesions and between their prevalence and rearing factors. *British Poultry Science 50*: 407-417. https://doi.org/10.1080/00071660903110901

Arnould C and Colin L 2008 Relation between a simple measure of activity, lameness and pododermatitis in meat type chickens. In: Koene P (ed) *Book of Abstracts of the 4th International Workshop on the Assessment of Animal Welfare at Farm and Group Level* p 73. 10-13 September 2008, Ghent, Belgium. Ponsen and Looijen: Wageningen, The Netherlands

Ask B 2010 Genetic variation of contact dermatitis in broilers. *Poultry Science 89*: 866-875. https://doi.org/10.3382/ps.2009-00496

Bilgili SF, Alley MA, Hess JB and Nagaraj M 2006 Influence of age and sex on footpad quality and yield in broiler chickens reared on low and high density diets. *The Journal of Applied Poultry Research 15*: 433-441. https://doi.org/10.1093/japr/15.3.433

Bruce DW, McIlroy SG and Goodall EA 1990 Epidemiology of a contact dermatitis of broilers. *Avian Pathology 19*: 532-537. https://doi.org/10.1080/03079459008418705

Bürkner PC 2017 brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software 80*: 1-28. https://doi.org/10.18637/jss.v080.i01

Byrt T, Bishop J and Carlin JB 1993 Bias, prevalence and kappa. *Journal of Clinical Epidemiology 46*: 423-429. https://doi.org/10.1016/0895-4356(93)90018-V

Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, Brubaker M, Guo J, Li P and Riddell A 2017 Stan: A probabilistic programming language. *Journal of Statistical Software 76*: https://doi.org/10.18637/jss.v076.i01

Ekstrand C, Algers B and Svedberg J 1997 Rearing conditions and foot-pad dermatitis in Swedish broiler chickens. *Preventive Veterinary Medicine 31*: 167-174. https://doi.org/10.1016/S0167-5877(96)01145-2

Ekstrand C, Carpenter TE, Andersson I and Algers B 1998 Prevalence and control of foot-pad dermatitis in broilers in Sweden. *British Poultry Science 39*: 318-324. https://doi.org/10.1080/00071669888845

European Commission Directorate-General for Health and Food Safety 2016 *Use of slaughterhouse data to monitor welfare of broilers on farm.* Luxembourg Publications Office of the European Union: Luxembourg

German Order on the Protection of Animals and the Keeping of Production Animals 2006 *Amended and promulgated on August 22nd, 2006, last changed on February 5th, 2014. Verordnung zum Schutz landwirtschaftlicher Nutztiere und anderer zur Erzeugung tierischer Produkte gehaltener Tiere bei ihrer Haltung; Tierschutz–Nutztierhaltungsverordnung (TierSchNutztV) in der Fassung der Bekanntmachung vom 22 August 2006 (BGBl. I S. 2043), die durch Artikel 1 der Verordnung vom 5 Februar 2014 (BGBl. I S. 94) geändert worden ist*

Greene JA, McCracken RM and Evans RT 1985 A contact dermatitis of broilers — clinical and pathological findings. *Avian Pathology 14*: 23-38. https://doi.org/10.1080/03079458508436205

Gunnarsson S, Algers B and Svedberg J 2000 Description and evaluation of a scoring system of clinical health in laying hens. In: Gunnarsson S (ed) *Laying Hens in Loose Housing Systems.* Doctoral Thesis, Swedish University of Agricultural Sciences, Uppsala, Sweden

Haslam SM, Knowles TG, Brown SN, Wilkins LJ, Kestin SC, Warriss PD and Nicol CJ 2007 Factors affecting the prevalence of foot pad dermatitis, hock burn and breast burn in broiler chicken. *British Poultry Science 48*: 264-275. https://doi.org/10.1080/00071660701371341

Heitmann S, Stracke J, Petersen H, Spindler B and Kemper N 2018 First approach validating a scoring system for foot-pad dermatitis in broiler chickens developed for application in practice. *Preventive Veterinary Medicine 154*: 63-70. https://doi.org/10.1016/j.prevetmed.2018.03.013

Hocking PM and Wu K 2013 Traditional and commercial turkeys show similar susceptibility to foot pad dermatitis and behavioural evidence of pain. *British Poultry Science 54*: 281-288. https://doi.org/10.1080/00071668.2013.781265

Implementary Rules of the Lower Saxonian Ministry of Nutrition 2015 *Agriculture and consumer protection for §20 (4; 5). Latest Edition: 31 July 2015, from the German Order on the Protection of Animals and the Keeping of Production Animals (2006).* http://www.nds-voris.de/jportal/portal/t/ji3/page/bsvorisprod.psml;jsessionid=8F93DE217C1A385FC1EBBE0F60477.jp27?pid=Dokumentanzeige&showdoccase=1&js_peid=Trefferliste&documentnumber=1&numberofresults=1&fromdoctodoc=yes&doc.id=VVND-VVND000035929#focuspoint

Landis JR and Koch G 1977 The measurement of observer agreement for categorical data. *Biometrics 33*: 159-174. https://doi.org/10.2307/2529310

Louton H, Erhard M and Wöhr A 2018 Acquisition of animal-based welfare measures at slaughter of poultry. *Fleischwirtschaft 11*: 94-98

Martland MF 1985 Ulcerative dermatitis in broiler chickens: The effects of wet litter. *Avian Pathology 14*: 353-364. https://doi.org/10.1080/03079458508436237

Martrenchar A, Boilletot E, Huonnic D and Pol F 2002 Risk factors for foot-pad dermatitis in chicken and turkey broilers in France. *Preventive Veterinary Medicine 52*: 213-226. https://doi.org/10.1016/S0167-5877(01)00259-8

Mayne RK, Hocking PM and Else RW 2006 Foot pad dermatitis develops at an early age in commercial turkeys. *British Poultry Science 47*: 36-42. https://doi.org/10.1080/00071660500475392

McKeegan D 2010 *Foot pad dermatitis and hock burn in broilers: Risk factors, aetiology and welfare consequences.* Research Project Final Report, Faculty of Veterinary Medicine, University of Glasgow, Scotland, UK. http://randd.defra.gov.uk/Default.aspx?Menu=Menu&Module=More&Location=None&Completed=0&ProjectID=14546

Meluzzi A, Fabbri C, Folegatti E and Sirri F 2008 Survey of chicken rearing conditions in Italy: Effects of litter quality and stocking density on productivity, foot dermatitis and carcase injuries. *British Poultry Science 49*: 257-264. https://doi.org/10.1080/00071660802094156

Michel V, Prampart E, Mirabito L, Allain V, Arnould C, Huonnic D, Le Bouquin S and Albaric O 2012 Histologically validated footpad dermatitis scoring system for use in chicken processing plants. *British Poultry Science 53*: 275-281. https://doi.org/10.1080/00071668.2012.695336

R Core Team 2018 *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing: Vienna, Austria. https://www.R-project.org/

Saraiva S, Saraiva C and Stilwell G 2016 Feather conditions and clinical scores as indicators of broilers welfare at the slaughterhouse. *Research in Veterinary Science 107*: 75-79. https://doi.org/10.1016/j.rvsc.2016.05.005

Shepherd EM and Fairchild BD 2010 Foot-pad dermatitis in poultry. *Poultry Science 89*: 2043-2051. https://doi.org/10.3382/ps.2010-00770

Spindler B and Hartung J 2011 Prevalence of pododermatitis in broiler chicken kept according to Directive 2007/43/EC Stocking Densities. *XVth ISAH Congress on Animal Hygiene International Society for Animal Hygiene (ISAH)* pp 39-42. 3-7 July 2011, Vienna, Austria

Weber Wyneken C, Sinclair A, Veldkamp T, Vinco LJ and Hocking PM 2015 Foot-pad dermatitis and pain assessment in turkey poults using analgesia and objective gait analysis. *British Poultry Science 56*: 522-530. https://doi.org/10.1080/00071668.2015.1077203

Welfare Quality® 2009 *Welfare Quality® Assessment Protocol for Poultry (Broilers, Laying Hens).* Welfare Quality® Consortium: Lelystad, The Netherlands

Wideman Jr RF 2016 Bacterial chondronecrosis with osteomyelitis and lameness in broilers: A review. *Poultry Science 95*: 325-344. https://doi.org/10.3382/ps/pev320

Wilcox CS, Patterson J and Cheng HW 2009 Use of thermography to screen for subclinical bumblefoot in poultry. *Poultry Science 88*: 1176-1180. https://doi.org/10.3382/ps.2008-00446

https://doi.org/10.7120/09627286.29.2.185 Published online by Cambridge University Press