



## OPEN ACCESS

## EDITED BY

Casey M. Owens,  
University of Arkansas, United States

## REVIEWED BY

Dianna Bourassa,  
Auburn University, United States  
Enver Cavusoglu,  
Bursa Uludağ University, Türkiye

## \*CORRESPONDENCE

Falko Kaufmann

✉ F.Kaufmann@hs-osnabrueck.de

RECEIVED 19 June 2024

ACCEPTED 23 August 2024

PUBLISHED 12 September 2024

## CITATION

Raederscheidt L, Kaufmann F, Spindler B,  
Kemper N and Andersson R (2024) Inter-  
observer reliability of a scoring system to  
evaluate bruises on turkey carcasses.  
*Front. Anim. Sci.* 5:1451488.  
doi: 10.3389/fanim.2024.1451488

## COPYRIGHT

© 2024 Raederscheidt, Kaufmann, Spindler,  
Kemper and Andersson. This is an open-access  
article distributed under the terms of the  
[Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).  
The use, distribution or reproduction in other  
forums is permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original publication in  
this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# Inter-observer reliability of a scoring system to evaluate bruises on turkey carcasses

Luisa Raederscheidt<sup>1,2</sup>, Falko Kaufmann<sup>1\*</sup>, Birgit Spindler<sup>2</sup>,  
Nicole Kemper<sup>2</sup> and Robby Andersson<sup>1</sup>

<sup>1</sup>Department of Animal Husbandry and Poultry Sciences, Osnabrueck University of Applied Sciences, Osnabrueck, Germany, <sup>2</sup>Institute for Animal Hygiene, Animal Welfare and Farm Animal Behavior, University of Veterinary Medicine Hannover, Foundation, Hannover, Germany

**Introduction:** Traumatic injuries such as bruises have been considered an important indicator to assess animal welfare in livestock farming. The possibility of assigning the injury to a particular stage or moment in the production process may allow judgments on possible causes and thus reduce its prevalence. Currently, there is no consistent definition or scoring system for bruised poultry carcasses in German meat inspection and the prevalence is affected by the variability of scoring systems as well as observer bias. Therefore, the objective of this study was to determine the inter-observer reliability of bruise characteristics at the slaughter line and to validate the findings with measurements of bruises and photographed carcasses.

**Methods & results:** Inter-observer reliability was assessed with two observers who each scored 10,880 turkey carcasses simultaneously at a running slaughter line after a short training session. The strength of agreement was “good” for the total number of detected bruises and the number of bruises on breasts, wings, and legs per flock (ICC = 0.81 – 0.88). Agreement (ICC) on the number of small, medium, and large bruises ranged between “good” and “moderate” values (0.70 – 0.84), whereas the number of bruises in different colors showed “moderate to poor” reliability (0.04 – 0.64). Additionally, agreement on bruise characteristics was tested using photographs (n= 513 carcasses) without a time limit. The highest agreement between observers was found for the location of bruises (Kw = 0.98). Again, the color of the bruises showed the lowest agreement (Kw = 0.36), whereas it was “moderate” for the size of the bruises (Kw = 0.45). When comparing each observer’s scoring values for size with size measurements (digital analysis) of the bruises, the observers tended to underestimate the actual size.

**Discussion & conclusion:** Overall, the total number of detected bruises and the location of the bruises showed the highest agreement between observers at the slaughter line and from photographs, indicating they were reliable variables. However, as the color variable showed a low agreement, a standardized objective method should be developed to objectively assess bruise prevalence and characteristics.

## KEYWORDS

bruises, animal welfare indicator, turkey carcasses, observer reliability, scoring system

## 1 Introduction

According to EU regulations, slaughterhouses are required to conduct animal welfare inspections (EU 2017/625, EU 2019/627). Traumatic injuries such as bruises are considered an important indicator in assessing animal welfare (Grilli et al., 2015; Huneau-Salaün et al., 2015; EFSA, 2019; Valkova et al., 2021). Despite the handling of living birds at the abattoir, there are a number of stages in the pre-slaughter chain that present a risk of injury to the birds, including rearing, catching, crating, and transport (Prescott et al., 2000; Nijdam et al., 2004; Allain et al., 2009; Gouveia et al., 2009; Langkabel et al., 2015; Jacobs et al., 2017a; Villarroel et al., 2018; Mönch et al., 2020). However, existing studies addressing the prevalence of bruises in poultry are very heterogeneous in their study design, and thus, the applied methodology and the obtained results vary (Prescott et al., 2000; Nijdam et al., 2005; Allain et al., 2009; Gouveia et al., 2009; Krautwald-Junghans and Felhaber, 2009; Allain et al., 2013; Jacobs et al., 2017b; Villarroel et al., 2018; Mönch et al., 2020). The main differences include the anatomical location assessed, the size threshold at which the bruises were examined, and the evaluation of the color of these bruises. Consequently, the prevalence of bruises on poultry carcasses is highly affected by the variability of the applied scoring systems, which makes it difficult to compare published results (Cockram and Dulal, 2018). Furthermore, visual scoring of lesions is influenced by observer bias (Meagher, 2009; Tuytens et al., 2014) and range of scale. Currently, there is no uniform definition or scoring system for bruised poultry carcasses in German meat inspection (Koch, 2016). Therefore, different examiners can gather different results when recording findings at the slaughter line (Hoischen-Taubner et al., 2011; Steinmann, 2018). However, the detection and characterization of bruises at the slaughter line may enable observers to attribute the injury to the production stage where it initially occurred, thereby allowing the implementation of precise measures to address its causes (Cockram and Dulal, 2018; EFSA, 2019).

Bruising is a trauma-induced injury (Barbut et al., 1990) without skin laceration (Northcutt et al., 2000) and extravasation of blood in the surrounding tissue (Langlois, 2007; Kostadinova-Petrova et al., 2017). The color change of bruises from red to violet, followed by shades of green and yellow due to degradation of hemoglobin, allows a rough chronological determination of the trauma event in poultry (Hamdy et al., 1961; Northcutt et al., 2000). However, perception of color is not only influenced by the subjective bias of observers but also by the age of the observer, external factors such as illumination, and the inconsistent color development of bruises (Stephenson and Bialas, 1996; Hughes et al., 2004, 2006; Grossman et al., 2011). Studies investigating the color description of bruising in humans in order to estimate its age have demonstrated low levels of agreement between observers (Munang et al., 2002; Bariciak et al., 2003; Pilling et al., 2010). Investigations on the reliability of a protocol for the evaluation of bruises in beef carcasses by Strappini et al. (Strappini et al., 2012). showed low inter-observer agreement as well. Some studies suggest combining different parameters of the injury characteristics such as location,

type, and age to draw conclusions regarding possible causes (Cockram and Dulal, 2018; Valkova et al., 2021). Therefore, it is necessary to objectify assessment methods in order to increase their reliability and thus, to use bruises as a suitable indicator.

The present study contributes to the ongoing debate by testing a developed scoring system at the slaughter line, on photographs, and on removed carcasses from the slaughter line. The findings of two observers were tested for agreement regarding bruise characteristics and the number of detected bruises. In order to evaluate the scoring scheme, the results of each parameter were additionally compared to data obtained with a more objective method. This involved comparing the numbers of observed bruises at the slaughter line and their localization to findings from photographs and scoring values for size and color to measurements. The objective of developing a scoring system for the systematic recording of bruises in turkeys at the slaughterhouse and determining the inter-observer reliability is to contribute to the development and implementation of a uniform and comparable survey method.

## 2 Materials and methods

Data were collected between September 2022 and March 2023 at a commercial turkey abattoir in Germany. Two observers evaluated bruises on turkey carcasses with a developed scoring system at the slaughter line, in photographs, and on carcasses removed from the slaughter line. The two observers were veterinarians, with one having over 20 years of experience in meat grading and the other a relatively inexperienced professional. Prior to the assessments, the observers underwent a brief training on the protocol and its scoring values via photographs and conducted a joint observation at the slaughter line. The training did not include a statistical test of observer agreement. The same two observers performed all observations during the trial.

### 2.1 Visual scoring system

The scoring system was developed based on literature research (Hamdy et al., 1961; Northcutt et al., 2000; Prescott et al., 2000; Krautwald-Junghans and Felhaber, 2009; Allain et al., 2013). Only bruises equal to or greater than 1 cm in diameter (measured at the largest extension) and bruises that were visible to the observers on the ventral side of the carcass were scored. Each detected bruise was assessed regarding the following parameters: location (limited to breast, wings, and legs); size classified as small (1 – 3 cm), medium (3 – 5 cm), or large (> 5 cm); and the color of the bruise, being either red, violet, green-violet, green-yellow, or yellow-orange based on Hamdy et al. (Hamdy et al., 1961) and Northcutt et al. (Northcutt et al., 2000) (Figure 1).

In general, the same scoring system was used for the visual assessment of bruises at the slaughter line, in photographs, and on removed carcasses.

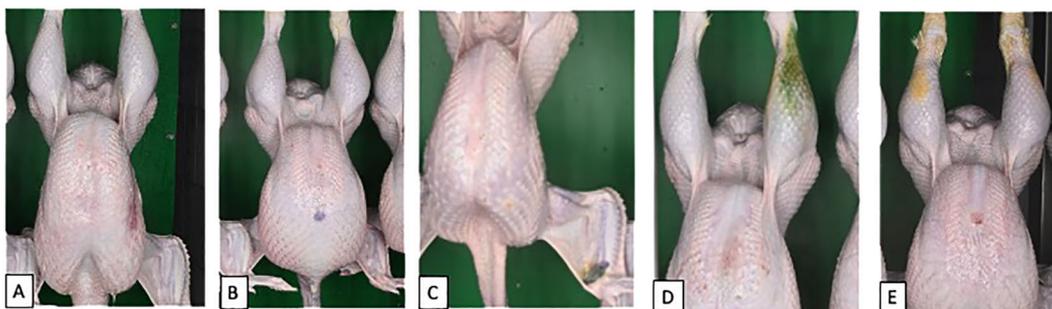


FIGURE 1

Scoring examples of bruised turkey carcasses (A) Large, red bruise on a breast; (B) Small, violet bruise on a breast; (C) Medium, green-violet bruise on a wing; (D) Large, green-yellow bruise on a leg; (E) Large, yellow-orange bruise on a leg.

## 2.2 Assessment at the slaughter line

Data were obtained from 20 randomly selected flocks at the slaughter line. Flock size differed between 1,400 and 10,000 birds and was on average 5,301 birds. In total, 10% of the birds of each flock (mean 530 birds) were sampled using the described scoring system. Of the 20 flocks, 10 male (average age: 20 weeks of life), six female (average age: 15 weeks of life), and four female breeder (average age: 60 weeks of life) flocks were evaluated. The turkey carcasses were assessed by the two observers after plucking and before evisceration. The scoring was carried out simultaneously but independently by the two observers. A grand total of 10,880 carcasses were examined at the slaughter line, with the slaughter line speed and thus, time for sampling, varying between 0.9 to 1.2 birds per second. As the carcasses were not individually marked, observer agreement was determined by the number of observed bruises and their characteristics per flock. A flock was defined as a group of turkeys that were housed together, raised in identical conditions, and depopulated and slaughtered at the same time (Allain et al., 2013).

## 2.3 Assessment via photographs

Standardized photographs of the ventral side of the carcasses were taken from 18 of the 20 flocks observed at the slaughter line with a Nikon Z7 camera (Melville, New York, USA). Independent of the observations at the slaughter line, 10% of the carcasses of each flock were randomly sampled (in total 8,690 carcasses) and later scored by one observer using the described scoring method. This was done to compare direct observations at the slaughter line with findings from the photographs, as the short observation time at the running slaughter line is often cited as a limiting factor in the evaluation of carcasses (Strappini et al., 2012; Törmä et al., 2021).

Additionally, 513 photographed turkey carcasses were randomly selected from the pool of photographs. The carcasses were numbered and independently examined by the two observers without any time limit. All documented bruises were measured (cm) using ImageJ Software 1.51j8 (National Institutes of Health, USA). For reference purposes, each carcass was photographed while

hanging on its shackle with a standardized width of 19.5 cm. Measurements were taken at the largest diameter of the bruise and were repeated three times by the same observer. Mean values were used for further analysis.

## 2.4 Assessment of retained carcasses

Randomly selected bruised carcasses were removed from the slaughter line and retained for further and detailed inspection. In total, 255 bruises were scored by the two observers using the visual scoring scheme. Furthermore, the color of each bruise was measured using a Minolta Chroma-meter<sup>®</sup> CR-400 (Osaka, Japan) Colorimeter. The Commission Internationale de l'Éclairage  $L^*a^*b^*$  (CIE- $L^*a^*b^*$ ) color space was selected, where the amount of white ( $L^* = 100$ ), black ( $L^* = 0$ ), red ( $+a^*$ ) green ( $-a^*$ ), yellow ( $+b^*$ ), and blue ( $-b^*$ ) is visualized and measured on three axes. Each bruise was measured at two different locations/sites: the center of the bruise and the margin of the bruise (0.5 cm into the visible bruise). Each measurement was recorded as the mean of three scans. The colorimeter was calibrated before measuring each bruise using a white calibration plate.

## 2.5 Statistical analysis

Inter-observer reliability between the two observers at the slaughter line was analyzed using the intra-class correlation coefficient (ICC) (3, 1) (Shrout and Fleiss, 1979) for metric data. Statistical analysis was performed using IBM<sup>®</sup> SPSS<sup>®</sup> Statistic Version: 28.0.1.0 (142) (IBM Corp, Armonk, New York, USA). The data were assessed for inter-observer reliability regarding the total number of observed bruises and the number of bruises according to location (breast, wings, and legs), size (small, medium, and large), and color.

To compare the findings from the slaughter line with the photographs, the ICC (3,1) was calculated between the observers and the number of bruises scored from the photographs.

The strength of agreement for ICC was evaluated according to Koo and Li (Koo and Li, 2016) as follows: Values less than 0.5

indicate poor reliability, values between 0.5 and 0.75 indicate moderate reliability, values between 0.75 and 0.9 indicate good reliability, and values greater than 0.90 indicate excellent reliability.

In order to evaluate the observer agreement regarding the bruise characteristics (ordinal variables) from photographs, weighted kappa (Kw) (Cohen, 1968) was calculated and statistical analysis was performed using SAS Version 9.4 (SAS Institute Inc., Cary, North Carolina, USA).

For the variable “observed bruises”, all 513 carcasses were considered. For the bruise characteristics “location”, “size”, and “color”, only bruises identified by both observers were evaluated (n = 124 bruises). A comparison of the results of visual scoring of size and measured bruise size was performed descriptively.

The strength of agreement for Kw was evaluated according to Landis and Koch (Landis and Koch, 1977). Values less than 0.0 indicate poor reliability, values between 0.0 and 0.2 indicate slight reliability, values between 0.21 and 0.4 indicate fair reliability, values between 0.41 and 0.6 indicate moderate reliability, values between 0.61 and 0.8 indicate substantial reliability, and values greater than 0.81 indicate almost perfect reliability.

A comparison of the scored colors and the measured color data (center and edge) between the observers for the retained carcasses was performed descriptively. Significant differences per observer were calculated using the paired t-test (level of significance: 0.05).

## 3 Results

### 3.1 Slaughter line

Each flock had an average of 530 carcasses evaluated (n = 20 flocks). The average number of scored bruises per flock was 69.3 (SD: 26.35) for Observer (OB) 1 and 46.9 (SD: 19.82) for OB 2, respectively (Table 1). Most of the observed bruises were located on the breasts and wings of the carcasses. Observers 1 and 2 detected an average of 9.6 (SD: 7.29) and 6.5 (SD: 4.87) large bruises per flock, respectively. However, both observers identified a considerably higher number of medium and small size bruises (Table 1). Observer 1 characterized the majority of the bruises as violet in color (51.5; SD: 21.76), whereas most of the bruises were characterized as red by OB 2 (41.8; SD: 20.18).

The ICC for the total number of observed bruises (ICC = 0.81) and location of the bruises (ICC range: 0.81- 0.88) showed a “good” agreement between the two raters. The ICC for the bruise characteristic “size” ranged between “moderate” agreement for the number of detected medium (ICC = 0.70) and large (ICC = 0.51) bruises and “good” agreement for the number of small bruises (ICC = 0.84). Regarding “color”, the agreement between observers was “poor” (ICC range: 0.04 - 0.45) except for green-violet bruises, where the agreement was “moderate” (ICC = 0.64).

For OB 1, the level of agreement between the assessment at the slaughter line and evaluation based on photographs was “good” for the total number of bruises (ICC = 0.82) as well as the number of scored bruises on breasts (ICC = 0.89) and wings (ICC = 0.77) (Table 2), whereas agreement on the number of bruises on legs was “moderate”. For OB 2, the ICC between the slaughter line and

photographs was “moderate” for the total number of bruises (ICC = 0.62) and the number of bruises on wings (ICC = 0.73) and legs (ICC = 0.56), whereas a “good” agreement was found for the number of observed bruises on breasts (ICC = 0.77).

### 3.2 Photographs

Weighted kappa values for characteristics of observed bruises from photographs are shown in Table 3. The highest agreement between the observers was found for the location of the bruises (Kw = 0.98). The color of the bruises was the variable with the lowest agreement (Kw = 0.36). The size of the bruises showed “moderate” reliability estimates between the observers (Kw = 0.45). The mean measured size of the scored bruises was 6.2 cm (median = 4.5 cm, min = 0.8 cm, max = 41.6 cm). Bruises between 1 and 3 cm (score 1, small) were all, except for one bruise, scored correctly by both observers (Table 4). Bruises between 3 and 5 cm (score 2, medium) were mostly evaluated as small (OB 1: 32 of 51 bruises; OB 2: 28 of 39 bruises). Using the visual scoring system, OB 1 classified 53 bruises as large, whereas 76 out of 171 bruises were verified as large (> 5 cm) after measurement. Observer 2 visually assessed only 13 of the 68 bruises that were larger than 5 cm correctly as score 3. In comparison with the measured values, OB 1 correctly estimated 69.7% and OB 2 19.1% of the bruises (recorded by at least one observer) in terms of size respectively.

### 3.3 Retained carcasses

The level of agreement between observers regarding the color of bruises on the carcasses removed from the slaughter line was “substantial” (Kw = 0.73, 95% CI [0.68, 0.78]). Characteristics (color data in L\*a\*b\*- values) for each observer’s visually scored colors are given in Table 5. Mean values for measured red (+a) were highest in “red” (OB 1: 16.8; OB 2: 11.7) and “violet” (OB 1: 6.6; OB 2: 8.9), whereas values for measured green (a-) were highest in “green-yellow” (OB 1: -3.5; OB 2: -2.2). The highest +b values representing yellow were found in “yellow-orange” (OB 1: 14.8; OB 2: 15.3) and “green-yellow” (OB 1: 13.6; OB 2: 11.7), whereas values for measured blue (b-) were highest in “violet” characterized bruises. Significant differences were found between the measured color values of the scored colors by the observers in the same color categories for all colors except green-violet (Table 5).

## 4 Discussion

Traumatic injuries, such as bruises, have been considered an indicator to assess animal welfare (Grilli et al., 2015; Huneau-Salaün et al., 2015; EFSA, 2019; Valkova et al., 2021). A reliable scoring protocol for carcass evaluation is required to obtain significant and comparable results. The objective of this study was to determine the inter-observer reliability of bruise characteristics at the slaughter line and compare the findings with measurements of bruises and photographed carcasses.

TABLE 1 Descriptive statistics (mean, standard deviation (SD), minimum (Min), maximum (Max)), with the level of agreement (intra-class correlation coefficient (ICC)) and 95% confidence intervals at the slaughter line by the pair of observers for the total number of observed bruises (total); the number of bruises on breasts, wings, and legs; the number of small (1- 3 cm), medium (3- 5 cm), and large (> 5 cm) bruises; and the number of bruises with the color red, violet, green-violet, green-yellow, and yellow-orange; n = 20 turkey flocks (10,880 carcasses).

Variable	Observer	Mean (SD)	Min	Max	ICC (95% CI)	Strength of agreement
Total	1	69.3 (26.35)	30	120	0.81 (0.51-0.92)	Good
	2	46.9 (19.82)	18	88		
Breast	1	25.8 (20.01)	3	76	0.82 (0.55-0.93)	Good
	2	17.7 (13.15)	0	38		
Wings	1	25.0 (13.62)	0	50	0.81 (0.51-0.92)	Good
	2	22.9 (13.92)	2	50		
Legs	1	8.6 (8.95)	0	33	0.88 (0.69-0.95)	Good
	2	6.3 (12.87)	0	45		
Small	1	30.3 (13.96)	2	58	0.84 (0.58-0.94)	Good
	2	28.0 (13.42)	7	54		
Medium	1	29.5 (14.19)	6	68	0.70 (0.23-0.88)	Moderate
	2	12.4 (7.65)	1	29		
Large	1	9.6 (7.29)	1	25	0.51 (0.25-0.80)	Moderate
	2	6.5 (4.87)	0	18		
Red	1	7.4 (5.25)	2	23	0.13 (-1.85-0.55)	Poor
	2	41.8 (20.18)	12	86		
Violet	1	51.5 (21.76)	14	102	0.04 (-1.43-0.62)	Poor
	2	1.0 (1.61)	0	7		
Green- violet	1	3.4 (2.96)	0	9	0.64 (0.10-0.86)	Moderate
	2	1.2 (2.04)	0	8		
Green- yellow	1	5.6 (5.29)	0	20	0.42 (-0.47-0.77)	Poor
	2	2.6 (2.30)	0	8		
Yellow-orange	1	1.4 (2.23)	0	10	0.45 (-0.4-0.78)	Poor
	2	0.4 (0.75)	0	2		

#### 4.1 Total number of detected bruises

An increase in the number of bruises in a flock may be an indication of an animal welfare problem that should be investigated (Krautwald-Junghans and Felhaber, 2009; Grandin, 2017). In order

to assess the actual impact on animal welfare, good inter-observer reliability in detecting bruises at the slaughter line is required (Strappini et al., 2012; Grandin, 2017).

The agreement between the observers on the total number of recorded bruises at the moving slaughter line was “good” (ICC=

TABLE 2 Descriptive statistics for the total number of observed bruises (total) and the number of bruises on breasts, wings, and legs from photographs with the level of agreement (ICC) with each observer’s findings at the slaughter line; n = 18 flocks (8,690 carcasses).

Variable	Mean (SD)	Min	Max	ICC (95% CI)	
				Observer 1	Observer 2
Total	110.3 (49.98)	36	221	0.82 (0.51-0.93)	0.62 (-0.01-0.86)
Breast	41.5 (26.83)	4	87	0.89 (0.72-0.96)	0.77 (0.40-0.92)
Wings	48.3 (26.75)	7	106	0.77 (0.38-0.91)	0.73 (0.28-0.90)
Legs	20.5 (9.04)	9	47	0.53 (-0.25-0.83)	0.56 (-0.18-0.83)

**TABLE 3** Weighted kappa ( $K_w$ ) values for observed bruises (Score: 0 = no detected bruise, 1 = one detected bruise, 2 = > 1 detected bruises per carcass) and bruise characteristics: location (Score 1 = breast, 2 = wing, 3 = leg), size (Score 1= small (1-3 cm), 2 = medium (3-5cm), 3 = large (>5cm)), and color (Score 1= red, 2= violet, 3=violet-green, 4= green-yellow, 5=yellow-orange) as scored by the two observers from photographs.

Variable	$K_w$ (95% CI)	Strength of agreement
Observed bruises	0.75 (0.66 – 0.78)	Substantial
Location	0.98 (0.95 – 1.00)	Almost perfect
Color	0.36 (0.25 – 0.48)	Fair
Size	0.45 (0.34 – 0.55)	Moderate

n = 513 turkey carcasses, n=124 bruises.

0.81). In comparison with the photographs, the level of agreement between OB 1 at the slaughter line and evaluation from photographs was “good” (ICC= 0.82) for the total number of bruises, while OB 2 showed only “moderate” (ICC= 0.62) agreement with photographic evaluation. Considering the mean values of observed bruises, OB 2 missed noticeably more bruises at the slaughter line than OB 1 (OB 1:  $\bar{X}$  69.3 bruises/flock; OB 2:  $\bar{X}$  46.9 bruises/flock; photographs:  $\bar{X}$  110.28 bruises/flock). Therefore, both the observer and the data source influence the prevalence. Strappini et al (Strappini et al., 2012). relates the variation between examiners to the speed of the slaughter line, the difficulty of scoring while the carcass is moving, and the observer’s respective experience. In our study, however, the greater experience of Observer 2 was not found to be associated with an enhanced ability to detect bruises at the slaughter line. Regular training and observer calibration could therefore be beneficial for both newly trained and experienced meat graders.

Contrary to our expectations, the inter-observer reliability of the total number of recorded bruises from photographs was slightly lower with “substantial” agreement than that at the slaughter line, even though there was no time limit during the assessment. The age of the observers can also be an influencing factor in the detection of bruises. Dos Reis et al (Dos Reis et al., 2020). showed a decreased

capacity to evaluate radiographic images when visual acuity is reduced. A decline in vision leads to a reduced ability to detect anatomical and contrast differences. Age-related decline in visual performance is a result of normal aging (Andersen, 2012) and may have influenced the ability to detect smaller and/or faint bruises in the photographs of the current study. In contrast, Pilling et al (Pilling et al., 2010). reported no effect of observer age on performance in the visual assessment of bruises.

## 4.2 Location

The number of bruises on breasts, wings, and legs (ICC range: 0.81 – 0.88) showed “good” agreement between the two raters at the slaughter line. The level of agreement between observers and evaluation using photographs also showed “good” (ICC OB 1: 0.89; OB 2: 0.77) estimates for breast bruises but only “moderate” (ICC OB 1: 0.53; OB 2: 0.56) values for leg bruises. This indicates that good inter-observer agreement is no guarantee for correct data, as observers may be biased in the same direction (Tuytens et al., 2014). The number of bruises on the wings showed “good” agreement between OB 1 and photographs (ICC = 0.77) and “moderate” reliability for OB 2 (ICC = 0.73). The visibility of certain carcass regions probably had an influence on the results. Due to the tightly packed slaughter line, the wings of the birds overlap and often only one wing or only parts of the wings were visible to the examiner. Bruises on the ventral base of the wings were generally difficult to detect when inspected from the front. The position of the inspectors in relation to the carcass may also play a role. In this study, the breast region was at the height of the observer, while inspecting the legs and wings created an angle (legs upwards and the wings slightly downwards). On average, most bruises were observed on the breast (OB 1:  $\bar{X}$  25.8 bruises/flock; OB 2:  $\bar{X}$  17.7 bruises/flock) and the wings (OB 1:  $\bar{X}$  25.0 bruises/flock; OB 2:  $\bar{X}$  22.9 bruises/flock) and fewer on the legs (OB 1:  $\bar{X}$  8.6 bruises/flock; OB 2: 6.3 bruises/flock). Expectation bias can influence subjective scoring methods as a result of different experiences and predispositions (Foddai et al., 2012; Tuytens et al., 2014). Thus, the experience presumably led to an

**TABLE 4** Measured size (in score values) of visually classified bruises from photographs.

Visual score	Observer	Measured score			
		1 (1-3 cm)	2 (3-5 cm)	3 (>5 cm)	total
1	1	44	32	6	82
	2	26	28	18	72
2	1	0	18	17	35
	2	1	10	37	48
3	1	0	1	53	54
	2	0	1	13	14
total	1	44	51	76	171
	2	27	39	68	134

Observer 1, n= 171 bruises; Observer 2, n= 134 bruises.

TABLE 5 Descriptive statistics for measured bruise color data in Lab-values (white ( $L^* = 100$ ), black ( $L^* = 0$ ), red (+  $a^*$ ), green ( $-a^*$ ), yellow (+  $b^*$ ), and blue ( $-b^*$ )) of each observer's (OB) visually scored bruise colors.

	OB	Mean (SD)			Min			Max		
		L	a	b	L	a	b	L	a	B
Red	1	56.4 <sup>a</sup> (9.04)	16.8 <sup>a</sup> (6.86)	4.4 <sup>a</sup> (4.80)	35.4	0.5	-10.2	74.0	32.3	12.4
	2	57.8 <sup>a</sup> (8.99)	11.7 <sup>b</sup> (8.15)	0.4 <sup>b</sup> (6.51)	36.3	-0.6	-12.5	73.0	30.0	12.4
Violet	1	59.5 <sup>a</sup> (9.78)	6.6 <sup>a</sup> (5.62)	-3.2 <sup>a</sup> (5.24)	27.8	-2.1	-12.5	74.5	25.2	9.1
	2	58.6 <sup>a</sup> (10.52)	8.9 <sup>b</sup> (7.38)	-0.7 <sup>b</sup> (6.06)	27.8	-2.1	-12.5	74.6	32.3	11.9
Green-Violet	1	55.3 <sup>a</sup> (8.18)	0.9 <sup>a</sup> (5.83)	8.1 <sup>a</sup> (6.19)	35.3	-14.0	-10.1	75.9	18.4	23.1
	2	55.9 <sup>a</sup> (8.11)	1.7 <sup>a</sup> (5.60)	6.9 <sup>a</sup> (7.62)	35.3	-10.8	-11.7	71.0	18.4	23.1
Green-Yellow	1	62.1 <sup>a</sup> (7.78)	-3.5 <sup>a</sup> (3.94)	13.6 <sup>a</sup> (5.63)	39.0	-13.0	-11.7	77.1	5.3	25.8
	2	61.6 <sup>b</sup> (8.33)	-2.2 <sup>b</sup> (5.16)	11.7 <sup>b</sup> (5.71)	39.0	-14.0	-5.0	81.0	12.2	25.8
Yellow-Orange	1	69.4 <sup>a</sup> (5.13)	2.5 <sup>a</sup> (3.03)	14.8 <sup>a</sup> (6.36)	50.9	-5.1	1.2	81.0	10.9	32.7
	2	67.2 <sup>b</sup> (7.03)	1.5 <sup>b</sup> (3.59)	15.3 <sup>a</sup> (6.88)	43.5	-8.6	-5.9	80.4	10.9	32.7

<sup>a,b</sup>: Different letters in the same column of each color indicate significant differences ( $p \leq 0.05$ ; paired t-test).

Red: OB1 n=67, OB2 n=77; violet: OB1 n=135, OB2 n=123; green-violet: OB1 n=118, OB2 n=80; green-yellow: OB1 n=104, OB2 n=142; yellow-orange: OB1 n=82, OB2 n=84 measurements.

expectation to find bruises on these anatomical sites. Given the speed of the line and the angle of vision, it is likely that the legs were unconsciously not examined as intensely as the other two localizations. This assumption is also supported by the “almost perfect” ( $Kw = 0.98$ ) observer agreement on the location of bruises when scored without a time restriction using the photographs. In contrast to our study, Strappini et al (Strappini et al., 2012). found only a “slight” agreement ( $ICC = 0.35$ ) on the number of bruises scored per site between observers at the slaughter line. However, a higher number of anatomical sites was evaluated on the beef carcasses (butt, rump-loin, ribs, forequarter, back, pin, and hip) which might have made it more difficult to clearly differentiate between body regions and influenced the variation between examiners.

### 4.3 Size

The size of a bruise is influenced by the severity and type of trauma and the affected tissue structure (Taylor and Helbacka, 1968; Kranen et al., 2000). The  $ICC$  for size ranged between “moderate” agreement for the number of medium ( $ICC = 0.70$ ) and large ( $ICC = 0.51$ ) bruises and “good” agreement for the number of small bruises ( $ICC = 0.84$ ) observed at the slaughter line. Observation using photographs without a time limit did not improve inter-observer agreement (“moderate”,  $Kw = 0.45$ ). Strappini et al (Strappini et al., 2012). also reported a “moderate” agreement ( $K = 0.43 - 0.56$ ) for the size of bruises, with “size” being the variable with the highest agreement between observers in their study. The measured size of visually classified bruises showed that the majority of small bruises were correctly scored as small. OB 2 in particular underestimated the medium and large bruises in size (Table 4), correctly estimating the size of only 19.1% of the bruises (compared to 69.7% for OB 1). The differing abilities of observers to estimate bruise sizes may potentially influence recorded prevalence,

especially if the protocols only include bruises that are at least a certain size. The median of all scored bruises was 4.5 cm (min 0.84 cm, max 41.55 cm), which suggests that scoring values for size (small: 1 - 3 cm, medium: 3 - 5 cm, large: > 5 cm) largely coincided with the actual measured values. In order to increase the observer agreement and to identify size differences, the size ranges of the individual scoring values could be extended (in cm).

### 4.4 Color

The enzymatic degradation of extravascular hemoglobin results in a discoloration of the skin and may help date bruises (Hamdy et al., 1961; Northcutt et al., 2000). At the slaughter line, the agreement for color between observers was “poor” ( $ICC$  range: 0.04 - 0.45) except for green-violet where the agreement was “moderate” ( $ICC = 0.64$ ). The lowest value with almost no agreement between examiners was found for “red” ( $ICC = 0.13$ ) and “violet” ( $ICC = 0.04$ ). The mean values of the detected red and violet bruises for both observers showed distinct differences in the distribution between the colors. Observer 1 assigned an average of only 7.4 bruises per flock to red, whereas OB 2 found an average of 41.8 red bruises per flock. However, OB 1 scored 51.5 bruises per flock as violet and OB 2 only 1.0. The differentiation within a color (e.g., red and violet) might be much more difficult than between two contrasting colors (e.g., red and green) and may not be feasible at the slaughter line for observers (Knock and Carroll, 2019). However, regarding the genesis of the bruises, the distinction between (bright) red and dark red/violet can make a time difference of up to 24 hours. Red (bright) bruises indicate a trauma that occurred approximately 2 minutes to 12 hours ago whereas dark red/violet bruises indicate a trauma that occurred between 12 and 24 hours ago (Hamdy et al., 1961; Northcutt et al., 2000). As Cockram and Dulal (Cockram and Dulal, 2018) pointed

out, there is potential for injury during each stage of poultry production, especially while handling/catching the birds for transportation or pre-slaughter handling and stunning (Nijdam et al., 2005; Delezie et al., 2006; Kittelsen et al., 2015; Langkabel et al., 2015; Jacobs et al., 2017b; Mönch et al., 2020; Gerritzen et al., 2022). Typically, in Germany, those stages would all be carried out within 24 hours. Therefore, the sole assignment of the bruises to the combined color red/violet cannot rule out any production step as the cause of the injuries. In addition, Strappini et al. (Strappini et al., 2012), only used three scores for the color of the bruises (red, bluish or dark colored, and yellow-orange) and still achieved only “slight” ( $k = 0.16$ ) to “fair” ( $k = 0.39$ ) agreement between observers.

Color measurements of each rater’s scored colors showed a significant difference between the same color categories, except for “green- violet” which also had the highest observer agreement at the slaughter line. Observer 1 tended to categorize bruises as “red” that had a higher red (+a\*) and yellow (+b\*) value than OB 2. Furthermore, bruises rated as “violet” by OB 1 had a significantly higher blue (-b\*) and a lower red (+a\*) value than those of OB 2. Thus, the measurements indicate that Observer 2 tended to categorize most of the violet bruises as red. The bruises in the green-yellow color category were significantly greener (-a\*) and more yellow (+b\*) for OB 1 than for OB 2. Measurements of the bruises scored as yellow-orange showed no significant difference in the yellow axes (+b\*) between the observers, suggesting a good ability to detect yellow. Studies on inter-observer variability and accuracy of visual assessment of bruises in forensic medicine also showed low reliability and subjective color perception (Munang et al., 2002; Barciak et al., 2003; Hughes et al., 2004; Pilling et al., 2010). Munang et al. (Munang et al., 2002), reported a large degree of variation in how different observers described the color of a bruise. Additionally, when comparing each observer’s color descriptions of the same bruises *in vivo* and from photographs, only 31% of comparisons were in complete agreement. Pilling et al. (Pilling et al., 2010), evaluated the accuracy of bruise age estimates by forensic experts based on visual assessment and found considerable inter- and intra-observer variability as well. The study of Hughes et al. (Hughes et al., 2004), revealed variability in the perception threshold for yellow and a decline in the ability to perceive this color in a bruise with age. Overall, the authors of these studies concluded that age estimates of bruises based only on visual assessment of color are rather unreliable.

When scoring bruises from photographs, the color of the bruises again showed the lowest agreement (“fair”  $Kw = 0.36$ ) of all the observed characteristics. The direct inspection of the bruised carcasses removed from the slaughter line noticeably increased the observer agreement for color to a “substantial” agreement ( $Kw = 0.73$ ). This could highlight the influence of limited time at the slaughter line and reduced perceptibility of bruises in photographs on observer agreement for color compared to direct examination of stationary carcasses. A previous study found that age estimation of bruises based on photographs is less accurate than direct examination, mentioning unreliable color reproduction and loss of contours as possible reasons (Pilling et al., 2010).

Overall, the total number of detected bruises and the location of bruises showed the highest agreement between observers at the slaughter line and from photographs, followed by size. When comparing the slaughter line results of the observers with the findings from the photographs of the same flocks, the total number and location of bruises showed “moderate” to “good” agreement for both observers for the photographs, indicating that they are reliable variables. Agreement on size was moderate between observers, suggesting variability in the ability to estimate sizes accurately. In general, the size tended to be underestimated by both observers. Color was the variable with the lowest agreement between the observers at the slaughter line and from photographs. However, in this study, direct examination of unmoving carcasses significantly increased agreement on the color of the bruises but might not be feasible at a commercial level. Further research would be required to investigate whether a reduction in the number of possible scores (e.g., color), an increase in the minimum size of the recorded bruises, or a focus on certain anatomical sites would improve observer agreement. However, this could reduce the value of the overall assessment of the examinations. Additionally, it is important to acknowledge that the study was conducted with only two observers. To ensure the validity of the results, future studies should include a larger number of observers. The results of the study provide an initial impression of the inter-observer reliability of a bruise-scoring protocol for turkey carcasses. In particular, the study highlights the limiting factors due to the subjective perception of the observers, most clearly seen in relation to the color of the bruises. An objective, measurable, and repeatable inspection that is not influenced by the subjective assessment of the observer and the conditions on the slaughter line is probably best achieved with the implementation of a technically standardized method.

## 5 Conclusion

The applied scoring protocol was not reliable for all analyzed bruise characteristics. Therefore, further observer training and/or the implementation of technically standardized methods are needed in order to improve the quality and validity of the data, thus enabling conclusions to be drawn regarding the circumstances under which trauma may have induced the development of a bruise.

## Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

## Ethics statement

Since this study did not involve any living animals, no ethical approval was required. The samples were obtained from turkeys routinely slaughtered in a commercial slaughterhouse in Germany.

## Author contributions

LR: Formal analysis, Investigation, Methodology, Project administration, Visualization, Writing – original draft. FK: Methodology, Project administration, Validation, Writing – review & editing. BS: Conceptualization, Writing – review & editing. NK: Conceptualization, Writing – review & editing. RA: Conceptualization, Funding acquisition, Supervision, Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by the Gesellschaft zur Förderung des Tierwohls in der Nutztierhaltung mbH (ITW), Schwertberger Straße 14, 53177 Bonn, Germany. Info@initiative-tierwohl.de.

## Acknowledgments

We want to thank the official veterinarians and technicians in the participating slaughterhouse for their assistance in gathering

## References

- Allain, V., Huonnic, D., Rouina, M., and Michel, V. (2013). Prevalence of skin lesions in Turkeys at slaughter. *Br. Poultry Sci.* 54, 33–41. doi: 10.1080/00071668.2013.764397
- Allain, V., Mirabito, L., Arnould, C., Colas, M., Le Bouquin, S., Lupo, C., et al. (2009). Skin lesions in broiler chickens measured at the slaughterhouse: relationships between lesions and between their prevalence and rearing factors. *Br. Poultry Sci.* 50, 407–417. doi: 10.1080/00071660903110901
- Andersen, G. J. (2012). Aging and vision: changes in function and performance from optics to perception. *WIREs Cogn. Sci.* 3, 403–410. doi: 10.1002/wcs.1167
- Barbut, S., McEWEN, S. A., and Julian, R. J. (1990). Turkey downgrading: effect of truck cage location and unloading. *Poultry Sci.* 69, 1410–1413. doi: 10.3382/ps.0691410
- Barcziak, E. D., Plint, A. C., Gaboury, I., and Bennett, S. (2003). Dating of bruises in children: an assessment of physician accuracy. *Pediatrics* 112, 804–807. doi: 10.1542/peds.112.4.804
- Cockram, M. S., and Dulal, K. J. (2018). Injury and mortality in broilers during handling and transport to slaughter. *Can. J. Anim. Sci.* 98, 416–432. doi: 10.1139/cjas-2017-0076
- Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *psychol. Bull.* 70, 213–220. doi: 10.1037/h0026256
- Delezie, E., Lips, D., Lips, R., and Decuyper, E. (2006). Is the mechanisation of catching broilers a welfare improvement? *Anim. Welfare* 15, 141–147. doi: 10.1017/S0962728600030220
- Dos Reis, C. S., Soares, F., Bartoli, G., Dastan, K., Dhalmi, Z. S., Hussain, A., et al. (2020). Reduction of visual acuity decreases capacity to evaluate radiographic image quality. *Radiography* 26, S79–S87. doi: 10.1016/j.radi.2020.04.012
- EFSA Panel on animal health and welfare (AHAW) (2019). Slaughter of animals: poultry. *EFSA J.* 17, e05849. doi: 10.2903/j.efsa.2019.5849
- EU 2017/625 Regulation (EU) 2017/625 of the European Parliament and of the Council of 15 March 2017 on official controls and other official activities performed to ensure the application of food and feed law, rules on animal health and welfare, plant health and plant protection products. Available online at: <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32017R0625> (Accessed March 19, 2024).
- EU 2019/627 Commission Implementing Regulation (EU) 2019/627 of 15 March 2019 laying down uniform practical arrangements for the performance of official controls on products of animal origin intended for human consumption in accordance with Regulation (EU) 2017/625 of the European Parliament and of the Council and amending Commission Regulation (EC) No 2074/2005 as regards official controls.

data and for their hospitality during visits to the slaughterhouse. Special thanks go to Dr. Dieter Mischok for his valuable and dedicated support during data collection.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Available online at: <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32019R062> (Accessed March 19, 2024).

- Foddai, A., Green, L. E., Mason, S. A., and Kaler, J. (2012). Evaluating observer agreement of scoring systems for foot integrity and footrot lesions in sheep. *BMC Vet. Res.* 8, 65. doi: 10.1186/1746-6148-8-65
- Gerritzen, M., Shynkaruk, T., Buchynski, K., Schwean-Lardner, K., and Crowe, T. G. (2022). "Chapter 6: poultry," in *Preslaughter handling and slaughter of meat animals* (Wageningen: Wageningen Academic Publishers), 233–265. doi: 10.3920/978-90-8686-924-4\_6
- Gouveia, K. G., Vaz-Pires, P., and da Costa, P. M. (2009). Welfare assessment of broilers through examination of haematomas, foot-pad dermatitis, scratches and breast blisters at processing. *Anim. Welfare* 18, 43–48. doi: 10.1017/S0962728600000051
- Grandin, T. (2017). On-farm conditions that compromise animal welfare that can be monitored at the slaughter plant. *Meat Sci.* 132, 52–58. doi: 10.1016/j.meatsci.2017.05.004
- Grilli, C., Loschi, A. R., Rea, S., Stocchi, R., Leoni, L., and Conti, F. (2015). Welfare indicators during broiler slaughtering. *Br. Poultry Sci.* 56, 1–5. doi: 10.1080/00071668.2014.991274
- Grossman, S. E., Johnston, A., Vanezis, P., and Perrett, D. (2011). Can we assess the age of bruises? An attempt to develop an objective technique. *Med. Sci. Law* 51, 170–176. doi: 10.1258/msl.2011.010135
- Hamdy, M. K., May, K. N., Flanagan, W. P., and Powers, J. J. (1961). Determination of the age of bruises in chicken broilers. *Poultry Sci.* 40, 787–789. doi: 10.3382/ps.0400787
- Hoischen-Taubner, S., Werner, C., and Sundrum, A. (2011). Aussagegehalt von Schlachthofdaten zur Verbesserung der Tiergesundheit. In: *Es geht ums Ganze: Forschen im Dialog von Wissenschaft und Praxis* (Berlin: Verlag Dr. Köster). Available online at: <https://orprints.org/id/eprint/17591/> (Accessed May 13, 2024).
- Hughes, V. K., Ellis, P. S., and Langlois, N. E. I. (2004). The perception of yellow in bruises. *J. Clin. Forensic Med.* 11, 257–259. doi: 10.1016/j.jcfm.2004.01.007
- Hughes, V. K., Ellis, P. S., and Langlois, N. E. I. (2006). Alternative light source (polilight®) illumination with digital image analysis does not assist in determining the age of bruises. *Forensic Sci. Int.* 158, 104–107. doi: 10.1016/j.forsciint.2005.04.042
- Huneau-Salaün, A., Stärk, K. D. C., Mateus, A., Lupo, C., Lindberg, A., and Bouquinn-Leneveu, S. L. (2015). Contribution of Meat Inspection to the surveillance of poultry health and welfare in the European Union. *Epidemiol. Infection* 143, 2459–2472. doi: 10.1017/S0950268814003379

- Jacobs, L., Delezie, E., Duchateau, L., Goethals, K., and Tuytens, F. A. M. (2017a). Broiler chickens dead on arrival: associated risk factors and welfare indicators. *Poultry Sci.* 96, 259–265. doi: 10.3382/ps/pew353
- Jacobs, L., Delezie, E., Duchateau, L., Goethals, K., and Tuytens, F. A. M. (2017b). Impact of the separate pre-slaughter stages on broiler chicken welfare. *Poultry Sci.* 96, 266–273. doi: 10.3382/ps/pew361
- Kittelsen, K. E., Granquist, E. G., Vasdal, G., Tolo, E., and Moe, R. O. (2015). Effects of catching and transportation versus pre-slaughter handling at the abattoir on the prevalence of wing fractures in broilers. *Anim. Welfare* 24, 387–389. doi: 10.7120/09627286.24.4.387
- Knock, M., and Carroll, G. A. (2019). The potential of post-mortem carcass assessments in reflecting the welfare of beef and dairy cattle. *Animals* 9, 959. doi: 10.3390/ani9110959
- Koch, M. (2016). Neukonzeption der Schlachtier- und Fleischuntersuchungsstatistik. *Statistisches Bundesamt WISTA- Wirtschaft und Statistik*. 6, 74–84. Available online at: [https://www.destatis.de/DE/Methoden/WISTA-Wirtschaft-und-Statistik/2016/06/neukonzeption-schlachtier-fleischuntersuchungsstatistik-062016.pdf?\\_\\_blob=publicationFile](https://www.destatis.de/DE/Methoden/WISTA-Wirtschaft-und-Statistik/2016/06/neukonzeption-schlachtier-fleischuntersuchungsstatistik-062016.pdf?__blob=publicationFile).
- Koo, T. K., and Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropractic Med.* 15, 155–163. doi: 10.1016/j.jcm.2016.02.012
- Kostadinova-Petrova, I., Mitevska, E., and Janeska, B. (2017). Histological characteristics of bruises with different age. *Open Access Maced J. Med. Sci.* 5, 813–817. doi: 10.3889/oamjms.2017.207
- Kranen, R. W., Lambooy, E., Veerkamp, C. H., Van Kuppevelt, T. H., and Veerkamp, J. H. (2000). Histological characterization of hemorrhages in muscles of broiler chickens. *Poultry Sci.* 79, 110–116. doi: 10.1093/ps/79.1.110
- Krautwald-Junghans, M.-E., and Felhaber, K. (2009). *Abschlussbericht zum Forschungsauftrag 06HS015 "Indikatoren einer tiergerechten Mastputenhaltung"*. Available online at: [file:///C:/Users/raeders/Downloads/Abschlussbericht\\_zum\\_Forschungsauftrag\\_06HS015:Indikatoren\\_einer\\_tiergerechten\\_Mastputenhaltung\\_%20\(1\).pdf](file:///C:/Users/raeders/Downloads/Abschlussbericht_zum_Forschungsauftrag_06HS015:Indikatoren_einer_tiergerechten_Mastputenhaltung_%20(1).pdf) (Accessed May 13, 2024).
- Landis, J. R., and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174. doi: 10.2307/2529310
- Langkabel, N., Baumann, M. P. O., Feiler, A., Sanguankiat, A., and Fries, R. (2015). Influence of two catching methods on the occurrence of lesions in broilers. *Poultry Sci.* 94, 1735–1741. doi: 10.3382/ps/pev164
- Langlois, N. E. I. (2007). The science behind the quest to determine the age of bruises—a review of the English language literature. *Forens. Sci. Med. Pathol.* 3, 241–251. doi: 10.1007/s12024-007-9019-3
- Meagher, R. K. (2009). Observer ratings: Validity and value as a tool for animal welfare research. *Appl. Anim. Behav. Sci.* 119, 1–14. doi: 10.1016/j.applanim.2009.02.026
- Mönch, J., Rauch, E., Hartmannsgruber, S., Erhard, M., Wolff, I., Schmidt, P., et al. (2020). The welfare impacts of mechanical and manual broiler catching and of circumstances at loading under field conditions. *Poultry Sci.* 99, 5233–5251. doi: 10.1016/j.psj.2020.08.030
- Munang, L. A., Leonard, P. A., and Mok, J. Y. Q. (2002). Lack of agreement on colour description between clinicians examining childhood bruising. *J. Clin. Forensic Med.* 9, 171–174. doi: 10.1016/S1353-1131(02)00097-4
- Nijdam, E., Arens, P., Lambooy, E., Decuyper, E., and Stegeman, J. A. (2004). Factors influencing bruises and mortality of broilers during catching, transport, and lairage. *Poultry Sci.* 83, 1610–1615. doi: 10.1093/ps/83.9.1610
- Nijdam, E., Delezie, E., Lambooy, E., Nabuurs, M. J., Decuyper, E., and Stegeman, J. A. (2005). Comparison of bruises and mortality, stress parameters, and meat quality in manually and mechanically caught broilers. *Poultry Sci.* 84, 467–474. doi: 10.1093/ps/84.3.467
- Northcutt, J. K., Buhr, R. J., and Rowland, G. N. (2000). Relationship of broiler bruise age to appearance and tissue histological characteristics. *J. Appl. Poultry Res.* 9, 13–20. doi: 10.1093/japr/9.1.13
- Pilling, M. L., Vanezi, P., Perrett, D., and Johnston, A. (2010). Visual assessment of the timing of bruising by forensic experts. *J. Forensic Legal Med.* 17, 143–149. doi: 10.1016/j.jflm.2009.10.002
- Prescott, N. B., Berry, P. S., Haslam, S., and Tinker, D. B. (2000). Catching and crating Turkeys: effects on carcass damage, heart rate, and other welfare parameters. *J. Appl. Poultry Res.* 9, 424–432. doi: 10.1093/japr/9.3.424
- Shrout, P. E., and Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *psychol. Bull.* 86, 420–428. doi: 10.1037/0033-2909.86.2.420
- Steinmann, T. (2018). (Hannover (Germany: University of Veterinary Medicine Hannover). Available online at: [https://elib.tiho-hannover.de/receive/etd\\_mods\\_00000035](https://elib.tiho-hannover.de/receive/etd_mods_00000035) (Accessed May 13, 2024). dissertation.
- Stephenson, T., and Bialas, Y. (1996). Estimation of the age of bruising. *Arch. Dis. Childhood* 74, 53–55. doi: 10.1136/adc.74.1.53
- Strappini, A. C., Frankena, K., Metz, J. H. M., and Kemp, B. (2012). Intra- and inter-observer reliability of a protocol for post mortem evaluation of bruises in Chilean beef carcasses. *Livestock Sci.* 145, 271–274. doi: 10.1016/j.livsci.2011.12.014
- Taylor, M. H., and Helbacka, N. V. L. (1968). A study of various factors affecting the bruising of broilers I. *Poultry Sci.* 47, 1616–1623. doi: 10.3382/ps.0471616
- Törmä, K., Lundén, J., Kaukonen, E., Fredriksson-Ahoma, M., and Laukkanen-Ninios, R. (2021). Prerequisites of inspection conditions for uniform post-mortem inspection in broiler chicken slaughterhouses in Finland. *Food Control* 130, 108384. doi: 10.1016/j.foodcont.2021.108384
- Tuytens, F. A. M., de Graaf, S., Heerkens, J. L. T., Jacobs, L., Nalon, E., Ott, S., et al. (2014). Observer bias in animal behaviour research: can we believe what we score, if we score what we believe? *Anim. Behav.* 90, 273–280. doi: 10.1016/j.anbehav.2014.02.007
- Valkova, L., Voslarova, E., Vecerek, V., Dolezelova, P., Zavelova, V., and Weeks, C. (2021). Traumatic injuries detected during post-mortem slaughterhouse inspection as welfare indicators in poultry and rabbits. *Animals* 11, 2610. doi: 10.3390/ani11092610
- Villarreal, M., Francisco, I., Ibáñez, M., Novoa, M., Martínez-Guijarro, P., Méndez, J., et al. (2018). Rearing, bird type and pre-slaughter transport conditions of broilers II. Effect on foot-pad dermatitis and carcass quality. *Spanish J. Agric. Res.* 16, e0504. doi: 10.5424/sjar/2018162-12015