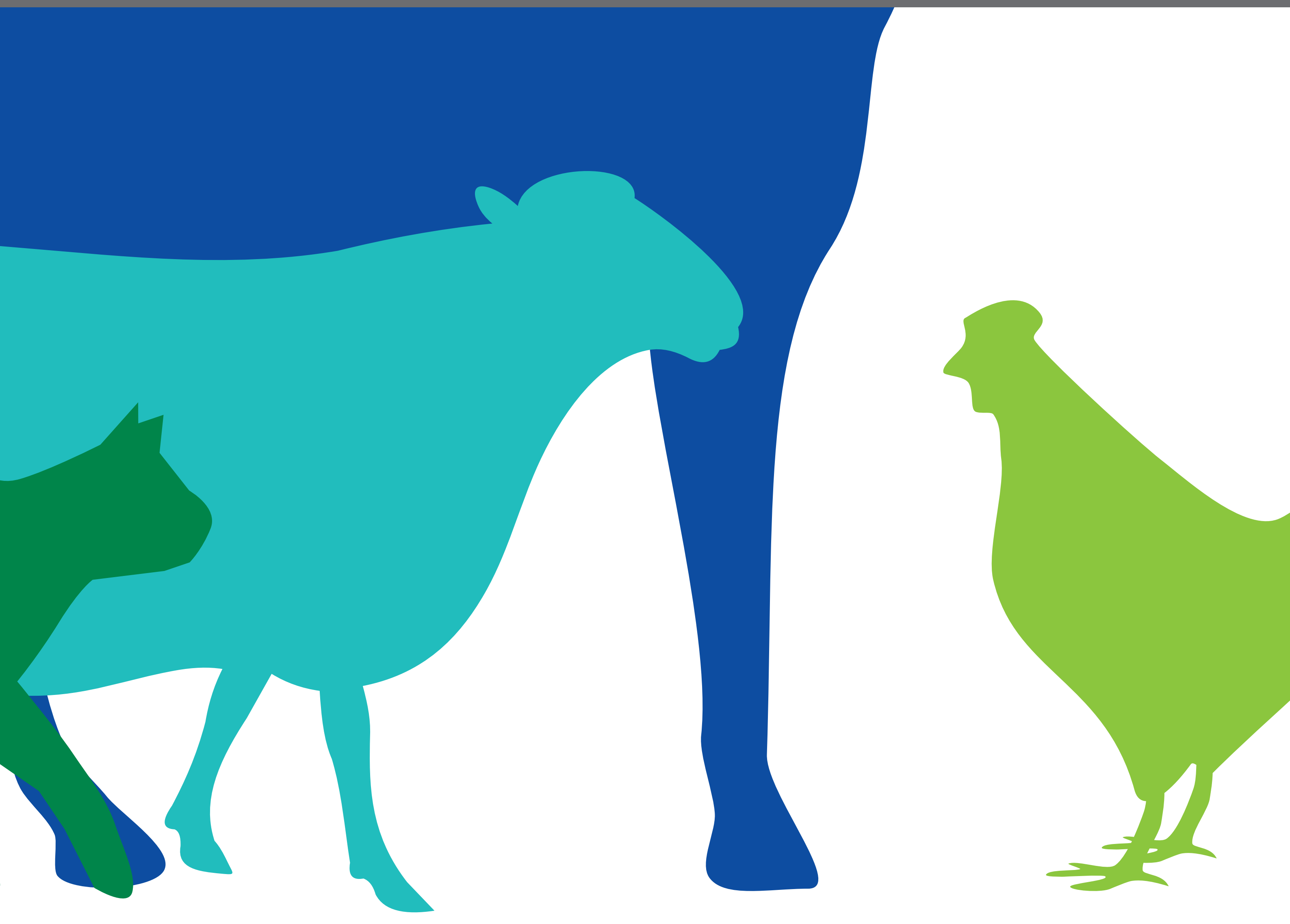
ANIMAL WELFARE ASSESSMENT: VOLUME 1

EDITED BY: Edward Narayan, Alan J. Tilbrook and Alan G. McElligott

PUBLISHED IN: Frontiers in Veterinary Science



frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88971-098-0
DOI 10.3389/978-2-88971-098-0

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

ANIMAL WELFARE ASSESSMENT: VOLUME 1

Topic Editors:
Edward Narayan, The University of Queensland, Australia
Alan J. Tilbrook, The University of Queensland, Australia
Alan G. McElligott, City University of Hong Kong, SAR China

Citation: Narayan, E., Tilbrook, A. J., McElligott, A. G., eds. (2021). Animal Welfare Assessment: Volume 1. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88971-098-0

Table of Contents

frontiers in Veterinary Science

EDITORIAL
published: 30 April 2021
doi: 10.3389/fvets.2021.653422

Editorial: Animal Welfare Assessment: Edition 1

Edward Narayan[1,2]*, Alan McElligott[3] and Alan Tilbrook[2]

[1] School of Agriculture and Food Sciences, Faculty of Science, The University of Queensland, St Lucia, QLD, Australia, [2] Centre for Animal Science, Queensland Alliance for Agriculture & Food Innovation, The University of Queensland, St Lucia, QLD, Australia, [3] Jockey Club College of Veterinary Medicine and Life Sciences, City University of Hong Kong, Hong Kong, China

Keywords: stress, animal welfare, livestock, wildlife, domesticated animals, biomarkers, climate change, affective state

Editorial on the Research Topic

Animal Welfare Assessment: Edition 1

Animal welfare refers to the well-being of the animal, and is the state of an animal as it tries to deal with the environment it is in. One way to assess the welfare state of the animal is through assessing whether it is under stress, which is the biological reaction when the animal is facing a potential threat to its welfare.

Although improvements in welfare such as through fine-scale adjustments to the animal's phenotype and its environment (e.g., genetics, husbandry, and nutrition) can improve the health of animals, good welfare does not always equate to increased productivity and vice-versa. The contextual nature of animal welfare issues transcend across different animal production systems whereby animals are managed by humans and require certain level of monitoring and care such as livestock, zoo animals, and pets. Productivity may be defined as a level of improved performance or fitness or a quantified production trait of an animal. For a farm animal, productivity could be a measured trait like meat production or milk quality while productivity for a pet species or zoo animal may not be as big a priority compared to animal welfare unless these animals are kept under breeding programs. Therefore, the current fragile atmosphere of shifting perspectives in the animal production sector and societal awareness in relation to the humane treatment of animals and use of animals for production has placed increasing pressure on finding a balance between management practices that can reduce stress, improve welfare and, equally, improve animal productivity.

In Edition 1 of this *Topic*, we show a collection of 12 peer reviewed articles which highlight the physiological, behavioral and physical health, and welfare evaluation of livestock and companion animals. It includes works of animal welfare experts, veterinarians, animal physiologists, and animal managers that will generate a healthy discussion and showcase latest studies working toward finding the harmony between animal production and welfare.

The papers presented in this special issue present new ideas and trialed research to boost animal health and welfare evaluation within intensive and extensive production systems as well as in pets and exotic species with examples from around the globe. For example, the first publication presented the physical and behavioral health indicators of cull cows in livestock markets. Sánchez-Hidalgo et al. developed a behavioral event index (BEI) comprising of cattle behaviors in the markets. Cow handler behavior was determined *via* negative tactile interactions (NTI) and the calculated index was termed as NTII. Researchers also evaluated the health status of each cattle. The researchers were successfully able to apply the cow and human related indices to determine the welfare of cull cows at livestock markets.

In the second animal welfare protocol-based research, Dalmau et al. presented a points-based animal welfare protocol for the farmed rabbits in Spanish farms by applying a multidimensional approach containing key animal-based indicators across age groups of rabbits.

OPEN ACCESS

Edited by:
Laura Ann Boyle,
Teagasc Food Research Centre, Ireland

Reviewed by:
Pol Llonch,
Universitat Autònoma de Barcelona, Spain
Susan Hazel,
University of Adelaide, Australia

***Correspondence:**
Edward Narayan
e.narayan@uq.edu.au

Specialty section:
This article was submitted to Animal Behavior and Welfare, a section of the journal Frontiers in Veterinary Science

Received: 14 January 2021
Accepted: 08 April 2021
Published: 30 April 2021

Citation:
Narayan E, McElligott A and Tilbrook A (2021) Editorial: Animal Welfare Assessment: Edition 1. Front. Vet. Sci. 8:653422. doi: 10.3389/fvets.2021.653422

In another of the works in this *Topic*, Teixeira et al., investigating animal-based welfare outcomes for pigs ($n = 54$ batches, 8,843 pigs) on-farm and abattoir in Chile, demonstrate that animal based physical health indicators can vary on farm and abattoir thus making it important to assess both individual farm and abattoir to enable through evaluation of pig welfare across the supply chain.

Tail biting is a significant welfare issue in intensive pig production. Haigh et al. applied open field and novel object test to test the animal related variation in stress responses to tail biting whereby preliminary results suggest that the differences in coping toward stress could be related to an individual pig's personality associated with either being a bold and tail biter or shy and victim.

Other contributions make relevant insights in applications of innovative tools for the animal welfare assessment. Magrin et al. studied a total of 2,161 animals from 80 Italian commercial farms. Researchers found specific lesions that could be input to develop a benchmarking system for evaluating animal health on-farm and applying this tool to improve the health and welfare of beef cattle.

In another paper, Kearton et al. show the application of associative learning behavior in Merino sheep to successfully train animals using classical conditioning to reduce contact with the aversive component of a virtual fence.

Three papers in this *Topic* focussed on companion animals, specifically dogs and cats. Davies, Scott et al., in their study successfully demonstrate the application of web based early warning system for providing 24/7 remote monitoring of dog well-being throughout the pet's lifetime. In the second dog-based paper, Clark et al. show the relationship between behavior and acute stress responses of both therapy dogs and their handlers. Salivary cortisol could be used to index stress levels of therapy dogs and applied in combination with behavior assessment to monitor the welfare of dogs. The paper by Davies, Reid, et al. was based on feline health evaluation using an online tool (HRQL). This tool helps to evaluate the impact of disease and clinical treatment on cat well-being and also supports clinical decisions and trials.

Production birds were also covered in this *Topic* with two papers. The team of researchers from Germany (Stracke et al.) inspected footpad dermatitis (FPD) in turkeys and show that improvements can be made to the current scoring system as a welfare tool through the evaluation of alterations on digits and using the total foot as a reference. In the second paper, Olschewsky et al. show the possibility of rearing slow growing turkey lines using organic husbandry which tends to improve health and welfare.

Finally, there is also a paper on fish welfare. Pedrazzani et al. show the application of an on-farm welfare assessment protocol for strengthening the practical application of on-farm welfare assessment in fish through the identification of critical welfare points.

Collectively, the *Topic* highlights current research areas and future directions in the dynamic field of animal welfare assessment. The variety of research papers demonstrate the availability of powerful tools in animal production systems through the combination of physiology, health and behavior indices and online monitoring systems to boost animal welfare for practical applications in research, commercial, and other settings.

AUTHOR CONTRIBUTIONS

EN conceptualized the Research Topic and collaborated with AM and AT for the coordination of this special issue. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by The University of Queensland, Faculty of Science, School of Agriculture and Food Sciences Start Up Funding awarded to EN.

ACKNOWLEDGMENTS

The editors thank the Frontiers admin team for their wonderful support with the special issue as well as the numerous peer reviewers.

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.

frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 07 August 2020
doi: 10.3389/fvets.2020.00471

Behavior and Health Indicators to Assess Cull Cow's Welfare in Livestock Markets

Melissa Sánchez-Hidalgo [1,2], Viviana Bravo [1,2] and Carmen Gallo [1]*

[1] Instituto de Ciencia Animal, Facultad de Ciencias Veterinarias, OIE Collaborating Centre for Animal Welfare and Livestock Production Systems—Chile, Universidad Austral de Chile, Valdivia, Chile, [2] Escuela de Graduados, Facultad de Ciencias Veterinarias, Universidad Austral de Chile, Valdivia, Chile

OPEN ACCESS

Edited by:
Edward Narayan,
The University of Queensland, Australia

Reviewed by:
Todd Duffield,
University of Guelph, Canada
Giuliana Miguel-Pacheco,
Independent Researcher, Nottingham, United Kingdom

***Correspondence:**
Carmen Gallo
cgallo@uach.cl

Specialty section:
This article was submitted to Animal Behavior and Welfare, a section of the journal Frontiers in Veterinary Science

Received: 24 April 2020
Accepted: 25 June 2020
Published: 07 August 2020

Citation:
Sánchez-Hidalgo M, Bravo V and Gallo C (2020) Behavior and Health Indicators to Assess Cull Cow's Welfare in Livestock Markets. Front. Vet. Sci. 7:471. doi: 10.3389/fvets.2020.00471

The welfare status of cull cows going through livestock markets was assessed in 12 premises in Chile, using behavioral and health indicators observed during unloading, auction, and loading (once in winter and once in summer). Groups of cows were observed by the same observer and the following indicators were recorded at each stage: slips, falls, balks, turns, jumps, and vocalizations of cows were considered as behavioral events and a proportion was calculated based on the number of observed events per group divided by the number of cows per group to give a behavioral event index (BEI). Health status of the cows was assessed during auction by recording the proportion of cows with low body condition, lameness, udder problems, tegumentary lesions, and tail abnormalities. Handler behavior was assessed using a count of negative tactile interactions (NTI) with the cows, like blows/hits, kicks, and pokes with the devices used to drive them, then a NTI index (NTII) was calculated as for BEI. Using the groups of cows as the statistical unit, statistical models were built and used to identify how NTII, some facilities features and comingling were associated with BEI registered during unloading, auction, and loading and also to calculate if the selling price was associated with the different health problems of cows, using the software MLwiN 3.03. A total of 1,103 groups of cows ($n = 3,963$ cows) were observed, finding a high percentage of slips and balks, whereas hitting and poking were frequent NTI. The highest mean BEI was observed during unloading in winter (1.10), whilst the lowest one was found during auction in summer (0.34). There was an increase of 0.11 in BEI for every extra unit increase in the NTII by the handlers. The BEI was negatively affected by the winter season compared to summer. Of 1,608 cows, 49.8% had a low body condition, 28.3% had udder problems, 24% were lame, 8.7% presented tegumentary lesions, and 3.1% tail abnormalities. It can be concluded that the health of the cull cows is already compromised when leaving the farms; cow behavior and handler tactile interactions with the cows are useful indicators to assess the welfare of cull cows at livestock markets.

Keywords: cull cows, livestock market, welfare indicators, behavior, handling, health

INTRODUCTION

Livestock markets are a traditional scenario to sell and buy cattle, and in Latin American countries are an important way for small farmers to sell their products (1). However, compared to selling directly from farm to farm or from farm to slaughterhouse, selling through auction markets implies greater concerns in terms of risks of transmission of diseases (2), increase in carcass bruising (3), and possibly greater animal welfare problems, considering that handling events during loading, transport, and unloading are at least duplicated and can generate high levels of stress (4).

Gallo and Tadich (5) indicate that comingling animals from diverse origins as it occurs at livestock markets implies an additional physiological and behavioral stress for the animals that can cause fatigue, fear, dehydration, hunger, weight loss, and lesions. It is well-known that the prevalence of bruises in carcasses of cattle that have undergone auction commercialization is higher than in carcasses of cattle sold directly from farms to slaughterhouses (3, 6, 7). The increased prevalence of bruises can be attributable to the fact that cattle sold through livestock markets undergo more handling events and have more human-animal, animal-animal, and also animal-facilities interactions (3, 8).

In Canada, sick dairy cows are an important welfare problem identified at livestock markets, particularly when the milk quota was reduced (4). In the United Kingdom the main welfare problem of livestock markets has been related to handling (6) and to infrastructure (9). In Chile De Vries (10) and Sepúlveda (8) found that in general, inadequate infrastructure and lack of trained personnel were relevant features at markets. Studies in livestock markets in the central region of Mexico (1, 2), have shown that animals present severe physiological, metabolic, and behavioral changes. In a recent study Bravo et al. (11) registered the behavior of weaned calves sold through livestock markets, as observed at the stages of unloading, grading, auction, loading, and penning and found that the main factors associated with poor welfare of the calves, according to their behavioral indicators, were related to bad handling techniques, infrastructure deficiencies, and lack of training of personnel.

Adult and old cows that are culled from beef and dairy farms represent between 30 and 40% of all animals sold through livestock markets in Chile (12). The physical and physiological conditions of culled cows differ widely from the younger and healthier cattle categories (13). Moreover, due to their comparatively lower commercial value, culled cows are usually handled with less care, are kept longer in lairage pens (14, 15), and show more bruises on their carcasses compared to steers and heifers (16, 17). At arrival at the slaughterhouse, the presence of health problems, a low body condition, and udder problems of culled cows also act as risk factors for the severity of bruises and carcass condemnations (18), increasing economic losses in the meat industry.

Although the effects of livestock markets on meat quality after slaughter are well-known, there is no sufficient scientific evidence regarding the effects of marketing on direct welfare indicators of the cows while alive. Therefore, the objective of the present study was to evaluate the welfare status of cows going through livestock markets in Chile, using behavioral, health, handler, and facilities indicators observed during the unloading, auction and loading process.

MATERIALS AND METHODS

The methodology used was similar to that described by Bravo et al. (11) in weaned calves and the evaluation guidelines were developed based on the recommendations of Welfare Quality (19) and Grandin (20) for slaughter plants, the current Chilean Animal Protection Law (21) and pilot visits to markets. Twelve auction markets in the southern regions of Chile (Geographical coordinates: 37.81208°S−72.67112°W to −45.61736°S−72.10496°W) were selected after pilot visits to the 21 existing southern markets; this selection was based on availability (auctioning weekly all year round or at least fortnightly) and on the number of cows arriving during 1 auction day (>50 cows). In order to evaluate possible seasonal differences, the markets were visited twice, the first visit was performed in summer and the second in winter. This considered mean daily temperatures (13.5 vs. 5.9°C) and pluviosity (31.6 vs. 194.7 mm) during summer (Nov-Dec) and winter months (June-July), respectively. The criterion to define cull cows at arrival at the market was a female bovine with some development of the udder, showing that she had had at least one calving, sometimes even accompanied by the calf.

The passage of the animals through the market was assessed in three stages, which are defined in **Table 1**. In each stage, groups of cows were observed by a veterinarian, trained in animal behavior, and welfare, in order to measure behavioral indicators of welfare during unloading, auction, and loading. Some other features related with facilities and handling by the personnel were also assessed independently in each group of cows observed at each market and stage; thus, animals were handled under comparable but not identical conditions at the different markets. To maintain consistency, the distinct stages were always assessed by the same observer throughout the study.

To assess the cow's welfare in each stage, the following indicators were evaluated.

Behavioral Indicators of the Cows

These were quantified in each observed group of cows, at each stage, in accordance with definitions provided by Maria et al. (22) and Gregory et al. (9), by counting slips, falls, balks, turns, jumps, and vocalizations. The number of individuals per group was also recorded. This included a count of other species involved during

TABLE 1 | Definition of the stages evaluated in livestock markets.

Stage	Description
Unloading	From the moment truck doors are opened until the last animal comes off the vehicle.
Auction	From cows entering the auction ring until leaving it.
Loading	From the lead animal moving toward the truck until the last animal is loaded into the truck.

unloading as necessary, as on occasion, mixed species groups could arrive at the market.

Slips

Foot slide or stumbling that did not result in a fall but nearly did so.

Falls

The cow went down on its side or both knees, or was off both its hind feet.

Balks

An unwanted arrested flow due to an apparent distraction or intimidation.

Turns

When the animal changes direction of movement against the animal flow.

Jumps

Leaping with all four feet simultaneously off the ground in a manner or situation that could be hazardous for the cow.

Vocalizations

When a cow makes a vocal sound (mooes).

Health Indicators of the Cows

The following health indicators were registered during auction, in accordance with definitions provided by Sánchez-Hidalgo et al. (18).

Body Condition

according to the European Welfare Quality (2009) protocol (19), a value of 0 was assigned to cows with a normal (regular) body condition, one to very lean cows (indicators for "very lean" present in at least three body regions) and two to very fat cows (indicators for "very fat" present in at least three body regions).

Abnormality of Tail

This was registered when the cows presented visually an amputated (shortened) or fractured tail (noticeable as an increase in volume and/or a lack of smooth continuity along the vertebrae when the tail was hanging relaxed). Fractured tails are common feature as a result of twisting tails by inappropriate handling.

Tegumentary Lesions

They were registered as such when the cows presented areas of alopecia in tarsus, hindquarters, carps, neck, shoulders, and back, as well as abrasions, scars, lacerations, hyperkeratosis, and other type of open wounds on their body.

Lameness

Category 0 without lameness was assigned to cows when timing of steps and weight-bearing was equal on all four feet. And one to cows with lameness (including categories 1 and 2 of the Welfare Quality Protocol, 2009) [i.e., if there was imperfect temporal rhythm in stride creating a limp or strong reluctance to bear weight on one limb, (19)].

Udder Problems

Cows presenting an increased volume of one or several mammary gland quarters, increased redness and inflammation of some quarters suggestive of mastitis, damaged, or visibly dry quarters, teats with visible wounds.

Additionally, during auction the price per kg live weight of the culled cows was registered in order to analyze the relationship between health indicators and price of the cows.

Handler Indicators

Handler-behavior was measured using a count of negative tactile interactions with the cows (blows/hits, kicks, and pokes with the devices used to drive the animals) during each stage, in accordance with definitions provided by Strappini et al. (3).

Blows/Hits

A person hits the animal with a hard object (usually wooden sticks).

Kicks

A person beats the animal with a foot.

Pokes

A person sticks a pointed object in the animal's body (usually pointed wooden sticks).

Additional Features

The facilities and comingling were also evaluated as possible features that could affect and/or disrupt the normal passage of the animals being driven.

Facilities

Floor type (slip-proof/slippery), obstacles/distractors in the path of the cows (presence or absence), and the slope of the ramp used (yes/no) (23) were recorded for each stage per group. Ramp slopes were obtained by simply measuring the height and the length of the ramp using a tape measure.

Comingling

In each stage it was recorded if cows were mixed with other categories of cattle, (e.g., heifers, bulls, steers, and calves) or with other animal species, (e.g., pigs, sheep, and horse) (yes/no).

Data Analysis

The slips, falls, balks, turns, and vocalizations were considered as behavioral events (BE). For each type of event, a proportion was calculated based on the number of observed events per group divided by the number of cows per group to give a behavioral event index (BEI). The hits, kicks and pokes by handlers with the device used to drive the animals, were classified as negative tactile interactions (NTI) and for each type of interaction, a proportion was calculated based on the number of observed interactions per group divided by the number of cows per group to give a NTI index (NTII). A descriptive analysis (mean, standard deviation (*SD*), minimum and maximum values and percentage calculation) was performed using IBM SPSS Statistics version 25, The body condition, abnormalities of tail, tegumentary lesions, lameness, and udder problems observed during the auction stage were counted as health indicators; one cow could have more than one health problem.

Statistical models were built and used to identify how NTII, some facilities features and comingling were associated with BEI as registered during unloading, auction, and loading. And also to calculate if the selling price was associated with the different health problems of cows. Multilevel model analyses were performed using the software MLwiN 3.03 (24). A multilevel modeling approach was employed to account for the clustering as a random effect and the repeated measurement structure of the data, (e.g., groups within auction market and repeated visits to market). Predictor variables were retained within models at $\alpha \leq 0.05$. A graphical inspection of the residuals was made to check for normality of errors and homogeneity of variance.

RESULTS

A total of 1,103 groups of cows ($n = 3,963$ cows) were observed. The size of the groups of cows ranged between one and 55, with a general mean of three. The descriptive analysis of the behavioral events observed in the cows in each stage is shown in **Table 2**. Slips and vocalizations were more often observed during auction, whereas the other behaviors registered were more frequent during unloading and loading; cows turning back were frequently observed during loading.

Regarding negative tactile human-animal interactions, animal handlers were often observed poking and hitting cows with driving devices, with higher frequency during loading; kicking cows was uncommon (**Table 3**).

Table 4 shows that the highest mean BEI was observed during unloading in winter (1.10), whilst the lowest one was found during auction in summer (0.34). The NTII was highest during loading, in particular in winter and lowest during auction in summer. For all events a minimum value of zero was found, meaning that in some groups no negative welfare events were registered.

Table 5 shows the results of the multinomial regression model used and the factors that had a significant effect on the BEI. The size of the group was significantly associated with the behavioral events, so that increasing by one the number of cows per group reduced the BEI in 0.03. There was an increase of 0.11 in BEI for every extra unit increase in NTII. The model shows that winter season increased BEI by 0.263 compared to summer. The presence of non-slippery floor at unloading and during auction decreased by 1.338 the BEI compared to cows handled on slippery floors. When cows were not mixed with other cattle categories (comingling), the BEI decreased by 0.551 compared with cows that were mixed.

Health Indicators of the Cows

Of 1,608 cows observed during auction, 49.8% ($n = 801$) had a very lean body condition, 28.3% ($n = 455$) had udder problems, 24% ($n = 386$) were lame, 8.7% ($n = 140$) presented tegumentary lesions, and 3.1% ($n = 50$) had abnormalities of the tail; all health problems were observed in higher percentages in winter than in summer. When exploring the relationship between price per kg live weight (Chilean pesos) and the health variables,

TABLE 2 | Descriptive analysis of behavioral events observed in the cows during unloading, auction and loading at the livestock markets.

Stage	Cows *N*	Behavioral events (%)					
		Slip	Fall	Balk	Turn	Jump	Vocalization
Unloading	1,580	13.7	4.1	11.4	0.5	3.0	7.7
Auction	1,608	15.5	1.0	6.3	1.6	0.9	10.0
Loading	775	9.3	3.6	10.3	15.6	1.2	4.0

TABLE 3 | Descriptive analysis of negative tactile interactions observed in the cows during unloading, auction and loading at the livestock markets.

Stage	Cows *N*	Negative tactile interactions (%)		
		Hits	Kicks	Pokes with the device
Unloading	1,580	13.2	4.6	16.3
Auction	1,608	22.6	0.0	6.8
Loading	775	53.9	1.3	85.3

TABLE 4 | Number of groups of cows observed, minimum, maximum, and mean (*SD*) indexes for behavioral events during movement at each stage of marketing per season.

Stage	Season	Groups	Behavioral events index (BEI)	
			Min-Max	Mean (*SD*)
Unloading	Winter	127	0–11	1.10 (1.87)
	Summer	130	0–18	0.75 (2.11)
Auction	Winter	385	0–9	0.57 (1.19)
	Summer	332	0–5	0.34 (0.63)
Loading	Winter	54	0–5	0.70 (1.09)
	Summer	75	0–10	0.68 (1.45)

TABLE 5 | Parameter estimate, standard error (*SE*), and significance for the models of average behavioral events index during unloading, penning, and loading of cull cows in the market.

Explanatory variables	Parameter estimate	S.E.	*z*-ratio	*p*-value
Constant	2.218	0.264	8.390	0.000
Group size	−0.037	0.009	−4.167	0.000
Negative tactile interactions	0.105	0.018	5.793	0.000
Season				
Winter	0.263	0.098	2.679	0.007
Summer	Reference			
Non slippery floor				
Yes	−1.338	0.262	−5.115	0.000
No	Reference			
Comingling				
No	−0.551	0.086	−6.412	0.000
Yes	Reference			

TABLE 6 | Parameter estimate, standard error (S.E.), and significance for the models of health and price per kg live weight during auction of cows in the market.

Explanatory variables	Parameter estimate	S.E.	*z*-ratio	*p*-value
	Price (Chilean pesos/kg live weight)			
Constant	828	21.4	38.660	< 0.001
Very lean body condition	−84	8.9	−9.475	< 0.001
Lameness	−23	10.4	−2.268	0.023
Tegumentary lesion	−77	15.8	−4.877	< 0.001

the multinomial regression model showed that a very lean body condition, lameness, and presence of tegumentary lesions affected negatively the selling price of the culled cows (**Table 6**).

The multinomial regression model indicates that cows with a very lean body condition had a 10% lower final price compared to cows with an appropriate (regular) body condition. The presence of lameness in cows reduced the selling price by 3% compared to not lame cows; the presence of tegumentary lesions reduced the price by 9% compared to cows without lesions. On the other hand, udder problems and tail abnormalities did not significantly affect the price of the cows, and therefore were not included in the model.

DISCUSSION

Results of the present study are based on the evaluation of behavioral and health indicators related to the welfare of the cows, as observed at the stages of unloading, in the auction ring, and loading at livestock markets. In general, it was found that cows presented a high proportion of negative behavioral events and health problems that reflect a poor welfare status, particularly during winter. The negative tactile interactions by the handlers were directly related to the behavioral events observed in the cows and the poor health status of the cull cows was also related to the price obtained for the cows. The results regarding negative tactile interactions by handlers and infrastructure problems when handling cull cows at livestock markets are similar to findings of earlier studies at this type of premises when observing different cattle categories in general (6, 9, 10) or just weaned calves (11). The welfare of cull cows has been assessed before at slaughterhouse level, at arrival, and during lairage, in Colombia (15, 25) and Chile (18). However, this is the first study trying to assess the welfare of the cows when passing through livestock markets using live animal-based indicators (health and behavior) as outputs and relating these to the handling and some other factors (facilities, comingling).

Behavioral Events of the Cows

Falls, slips, jumps, balking, turning, and vocalizations during driving of animals are behavioral events associated with fear and discomfort and their measurement can therefore be used to identify a welfare problem (11, 20, 26, 27).

According to Grandin (20) falls of animals during handling are acceptable up to 1%, slips and vocalizations of animals are acceptable up to 3%; an increase in these proportions reflects poor welfare. Our descriptive results for the behavioral events (**Table 2**) indicate that the presence of slips, falls, and vocalizations during unloading, auction and loading at cattle markets, were higher than recommended and possibly generate a decrease in the welfare of the cull cows during marketing. Our results for cows are in agreement with a study at 18 livestock markets in the United Kingdom, where Gregory et al. (9) also observed that the main welfare problems in fat cattle and calves marketed were slips and falls.

Turns and balks were more frequent during loading of the cows (**Table 2**). Moreover, negative tactile interactions by the handlers were also high at loading (**Table 3**). Both the high frequency of cow's turns and backs, and of NTI by handlers could reflect a problem with loading facilities, distractors at loading and also with the lighting and tiredness of the handlers, as it was observed that loading took place at the end of the marketing day and often occurred at night time.

Jumps were uncommon in the cows, whereas this behavior has been frequently observed in weaned calves during marketing (11). The results of the present study showed that some behavioral indicators (falls, slips, jumps, and vocalizations) were higher during unloading compared to loading, which suggests that this stage also generates discomfort and stress in the cows, probably due to the fact that they arrive at a new environment, where they are handled roughly by people. The aforementioned results, together with the health problems observed probably make cull cows more prone to lose balance, slip or fall, and hence decrease their welfare when passing through livestock markets.

Negative Tactile Interactions With Handlers

NTI like hitting, poking, kicking, and even tail twisting have been often observed during handling at cattle markets as it has been reported in earlier studies in the United Kingdom (4, 9); in Bangladesh (28); in Colombia (29, 30) and also in Chile (8, 10, 11). The most common NTI in the present study were hits and pokes, whereas kicking was uncommon (**Table 3**). The most common devices used in these interactions during loading and unloading were wooden sticks, sticks made of plastic pipe pieces and electric prods; during auction the most frequently driving device used was a plastic flap. It is worth mentioning that kicking, hitting and poking are banned by the Chilean regulation (23). Our finding is similar to an earlier study in Chile (8) where it was found that forbidden human-animal interactions were still present in 85.7% of the cattle markets evaluated. Considering that the regulations that prohibit these handlings were passed in 2013, more auditing by the competent authority is needed. The Chilean regulation (23) also includes compulsory training of the personnel handling the animals at cattle markets. Although in the view of the authors there have been improvements in several cattle markets compared to de Vries (10) and Sepúlveda (8), more supervision of the behavior of the personnel during the different stages of marketing is required by market managers (usually veterinarians) as well as by the competent authority.

Various authors recommend evaluating handling indicators because these trigger animal behaviors related to fear and escape, like slips, falls, jumps, balking, aggressions, and vocalizations

(31–34). These indicators also reflect the efficiency with which animals are handled by the personnel (11, 26, 27, 35). We found that most NTI occurred during loadings, particularly during winter, and the least occurred during auction in summer (**Table 4**). The loading of animals is the most critical point of those evaluated at the cattle markets, which had also been pointed out by Maria et al. (22). Considering this, the stage of loading animals would be a good one to observe for assessing welfare problems at livestock markets.

There was a significant direct relationship between the BEI and NTII, so that BEI increased when inappropriate handling like kicking, poking, and hitting increased. Our results are in accordance with an earlier study in calves in the UK, where the authors observed that 11% of calves fell during loading when they were inappropriately handled (9) and with a recent study of the behavior of calves at the same markets, in Chile (11). In order to reduce the impact of NTI on BEI and improve the welfare of animals at livestock markets, it is important that stockmen are trained and act according to the existing legislation.

Health Indicators

Health problems are obvious reasons for culling cows (36) and therefore it was not surprising to find a high proportion of cull cows with this type of problem at livestock markets. A study conducted in Chile determined that the main causes of elimination of dairy cows are reproductive problems, mammary gland affections, and lameness (37). Coincidently, our study showed that a high percentage of the cows arriving at the livestock markets had at least one health problem, mainly very lean body condition (49.8%), udder problems (28.3%), and lameness (24%). These results are also in agreement with studies of cull cows conducted in slaughterhouses in Colombia (25), in Sudan (38), and in Denmark (36). Similar results were obtained in a parallel study of cull cows arriving directly from farms to a slaughterhouse in Chile, where 52% of the cows presented one or more of the same health problems (18).

Considering the chronic characteristics of most of the health problems recorded in the cull cows, the fact that the cows were transported for a short time (∼4 h) directly from the farms of origin to the livestock markets and the short time spent in the different stages at the livestock market (∼12 h), the high prevalence of health problems most likely indicates that the cows were already sick or in poor body condition when leaving the farms of origin. Notwithstanding this, future research is needed at farm level in order to record the health problems present in cull cows before transport to the livestock markets or slaughterhouses and the cow's fitness for transport, as the clinical condition of them may deteriorate during transport (36). On the farm, cull cows with low body condition are at a greater risk of suffering of other illnesses (39). At slaughter, the low body condition is a risk factor for increasing the number of bruises on the carcass and for carcass condemnations (18), generating economic losses. That is why Grandin (20) and Losada et al. (27) recommend to cull cows before they are in such a bad state, when their welfare and health is severely compromised. Considering all the above mentioned, selling cull cows through cattle markets in these bad health conditions represents an additional effort and is negatively affecting their welfare during transport and marketing in comparison to younger and healthier animals like calves, steers, and heifers.

Culling the cows before they are in poor condition could also have economic benefits, because a higher price could be reached and/or the carcass will not be condemned after slaughter (18). With respect to the sale price in the market, we found a statistical association between the price per kg live weight and the health variables, where the body condition score, presence of lameness, and the tegumentary lesions negatively affected the selling price of the culled cows. Similar results were found in livestock markets in Canada, where cows in poor body condition or abnormal gait were sold at lower prices compared to cows in good body condition (40). From an animal welfare point of view and considering economic losses, it would be better to send culled cows directly from farms to slaughterhouses instead of passing through the cattle markets, particularly those in low body condition and with health problems.

Other Factors

Several authors mention that there are more slips and falls when the concrete floor is wet or dirty with urine and feces (9, 11, 31), which is in line with the results of the present study (**Table 5**), where the BEI increased in winter compared to summer, (i.e., when it was rainy). Facilities become slippery when wet and there is probably also an effect of the speed at which the handlers move the animals, trying to finish work quickly with bad weather conditions, as livestock markets in Chile lack roofing. We found that the lack of appropriate floor type was a predictor variable for an increase in the BEI. Often when the floor was described as slippery, we noticed that the surface was slippery due to wear through heavy use over the years. Similar results have been described by Gregory et al. (9); these authors mention that in order to solve this problem, some markets in the United Kingdom have put resin on the concrete floor, particularly on high risk areas, and have observed reductions in slipping. Considering that many cull cows arrive in poor body condition or lame, and have difficulties in moving, it is of primary importance to keep surfaces non-slippery.

Another important factor to be considered when moving the cows at livestock markets is the group size (**Table 5**); when cattle are separated from their original groups, they are more difficult to be handled (41). In this case we had small groups of cows (mean = 3), ranging between one and 55, and coincidently, we found a significant association between the size of the cow group and the BEI, so that increasing the number of cows in the group reduced the BEI. This result is similar to that found for weaned calves at the same livestock markets (11) and coincides with Grandin (41) that groups of cattle are easier to move than one animal alone. Mixing unfamiliar cattle from different sources or with other categories of cattle has often been observed at Chilean markets (10, 11). Comingling triggers agonistic behaviors related to dominance, like mounting and aggressions; further on this implies movement of animals that can end in cattle slipping, falling, balking, turning, and vocalizing (11, 41). In our study comingling cull cows with other sources and categories of cattle was associated with a higher BEI and hence should be avoided for improving welfare, besides of its effect on the transmission of diseases.

CONCLUSIONS

During marketing of cull cows, negative tactile interactions by handlers, like hitting, poking, and kicking were associated with an increase in behavioral events in the cull cows, like falling, slipping, jumping, turning, and backing. A small group size, winter season, presence of slippery floor, comingling with other cattle categories were also associated with an increase of behavioral events in the cows.

The high proportion of cull cows presenting low body condition and chronic health problems like lameness, udder problems, and others during auction reflect that there is a welfare problem not only at market level, but also on farm. Because sick and low body condition cows have more difficulties when they need to be moved, loaded, and unloaded several times, it is recommended that cull cows should be sold directly from farm to slaughterhouse and considering fitness for transport, in order to avoid further welfare problems and to reduce economic losses from this cattle category.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The animal study was reviewed and approved by Bioethics Committee Use of Animals in Research at Austral University of Chile (Application N325/2018).

AUTHOR CONTRIBUTIONS

MS-H and VB were responsible for the elaboration of behavior and handling protocols, general design of the study, and organization of the visits to the livestock markets. MS-H was responsible for the elaboration of health protocols, the collection of behavior, handling, and health data. CG was responsible for contacting livestock managers, general supervision, and acquisition of funding. VB did the statistical analyses and modeling of data. All authors contributed to the writing and discussion of the manuscript and approved its final version.

ACKNOWLEDGMENTS

The authors wish to acknowledge the collaboration of the Animal Welfare Group of Universidad Austral de Chile, particularly Silvia Larios, Mónica Lozano, Javiera Calderón, and Cristian Larrondo in the collection of data, as well as all the managers of the participating livestock markets who facilitated the visits. We also thank Escuela de Graduados, Facultad de Medicina Veterinaria, Universidad Austral de Chile for the economic support to accomplish this study.

REFERENCES

1. Roldan P, De la Cruz L, Tarazona A, Buenhombre J, Acerbi R, Varona E, et al. Bienestar animal en mercados ganaderos. In: Mota D, Velarde A, Huertas S, Cajiao M, editor. *Bienestar Animal: Una Visión Global En Iberoamérica*. España: Elsevier (2016). p. 155–68.
2. Corrales-Hernández A, Mota-Rojas D, Guerrero I, Roldan P, Rodríguez S, Yáñez A, et al. Physiological responses in horses, donkeys and mules sold at livestock markets. *Int J Vet Sci Med.* (2018) 6:97–102. doi: 10.1016/j.ijvsm.2018.03.002
3. Strappini A, Frankena K, Metz J, Gallo C, Kemp B. Characteristics of bruises in carcasses of cows sourced from farms or from livestock markets. *Animal.* (2012) 6:502–9. doi: 10.1017/S1751731111001698
4. Gregory NG. Animal welfare at markets and during transport and slaughter. *Meat Sci.* (2008) 80:2–11. doi: 10.1016/j.meatsci.2008.05.019
5. Gallo C, Tadich N. Bienestar animal y calidad de carne durante los manejos previos al faenamiento en bovinos. *REDVET*. (2008) 9:1695–7504. Available online at: https://www.redalyc.org/articulo.oa?id=636/63617111001
6. Weeks CA, McNally PW, Warriss PD. Influence of the design of facilities at auction markets and animal handling procedures on bruising in cattle. *Vet Rec.* (2002) 150:743–8. doi: 10.1136/vr.150.24.743
7. Edwards-Callaway LN, Walker J, Tucker CB. Culling decisions and dairy cattle welfare during transport to slaughter in the United States. *Front Vet Sci.* (2019) 5:343. doi: 10.3389/fvets.2018.00343
8. Sepúlveda CF. *Evaluación de las condiciones de bienestar de los animales comercializados en ferias ganaderas de la Novena Región de la Araucanía de acuerdo al cumplimiento del Decreto N° 29 de la Ley de Protección Animal* (Memoria de título). Temuco, Universidad Mayor Chile (2015).
9. Gregory NG, Benson T, Mason CW. Cattle handling and welfare standards in livestock markets in the UK. *J Agr Sci.* (2009) 147:345–54. doi: 10.1017/S0021859609008508
10. De Vries M. *Human-Animal relationship at Chilean livestock markets* (Ph.D. thesis). Wageningen University, Netherlands (2011).
11. Bravo V, Knowles TG, Gallo C. Factors affecting the welfare of calves in auction markets. *Animals.* (2019) 9:333. doi: 10.3390/ani9060333
12. ODEPA (Oficina de Estudios y Políticas Agrarias). *Boletín semanal de precios de la Asociación Gremial de Ferias Ganaderas*. (2018). Available online at: http://www.odepa.cl (accessed February 9, 2019).
13. Gallo C, Strappini A. Ensuring the welfare of culled dairy cows during transport and slaughter. In: Webster J, editor. *Achieving Sustainable Production of Milk Volume 3: Dairy Herd Management and Welfare*. Bristol, UK: Burleigh Dodds Science Publishing (2017). p. 103–22.
14. Carmine, XV. *Análisis del tiempo de transporte y espera, destare y rendimiento de la canal de bovinos transportados desde Osorno a Santiago* (Tesis de Licenciatura). Valdivia, Universidad Austral de Chile (1995).
15. Ramírez MF, Gallo C, Rosenfeld C. Bienestar animal durante el manejo de bovinos en una planta faenadora en Colombia: diagnóstico e intervención. *Revista ACOVEZ.* (2017) 46:9–12. Available online at: http://www.acovez.org/images/Revistas/Edicion-128.pdf
16. Sandoval M. *Estudio de las contusiones presentes en canales de bovinos procedentes de ferias y predios faenados en el frigorífico de Temuco* (Memoria de Título). Temuco, Universidad Católica de Temuco (2007).
17. Strappini A, Metz J, Gallo C, Kemp B. Origin and assessment of bruises in beef cattle at slaughter. *Animal.* (2009) 3:728–36. doi: 10.1017/S1751731109004091
18. Sánchez-Hidalgo M, Rosenfeld C, Gallo C. Associations between pre-slaughter and post-slaughter indicators of animal welfare in cull cows. *Animals.* (2019) 9:642. doi: 10.3390/ani9090642
19. Welfare Quality. *Welfare quality assessment protocol for cattle*. (2009). Available online at: http://www.welfarequalitynetwork.net/en-us/reports/assessment-protocols/ (accessed March 13, 2020).
20. Grandin T. Auditing animal welfare at slaughter plants. *Meat Sci.* (2010) 86:56–65. doi: 10.1016/j.meatsci.2010.04.022

21. Chile. *Ley 20.380 Sobre protección de los animales. Ministerio de Salud Pública, Chile*. (2009). Available online at: https://www.leychile.cl/Navegar?idNorma=1006858 (accessed March 3, 2020).
22. María GA, Villarroel M, Chacon G, Gebresenbet G. Scoring system for evaluating the stress to cattle of comercial loading and unloading. *Vet Rec.* (2004) 154:818–21. doi: 10.1136/vr.154.26.818
23. MINAGRI. *Reglamento sobre protección de los animales durante su producción industrial, su comercialización y en otros recintos de mantención de animales. Ministerio de Agricultura, Chile*. (2013). Available online at: https://www.leychile.cl/Navegar?idNorma=1051298 (accessed March 3, 2020).
24. Charlton C, Rasbash J, Browne WJ, Healy M, Cameron B. *MLwiN Version 3.03; Centre for Multilevel Modeling*. Bristol: University of Bristol (2019).
25. Romero M. *Estudio de indicadores fisiológicos de salud en vacas de descarte*. (Ph.D. dissertation). III Encuentro Internacional de Investigadores de Bienestar Animal, México (2015).
26. Hemsworth P, Rice M, Karlen M, Calleja L, Barnett J, Nash J, et al. Human-animal interactions at abattoirs: Relationships between handling and animal stress in sheep and cattle. *Appl Anim Beh Sci.* (2011) 135:24–33. doi: 10.1016/j.applanim.2011.09.007
27. Losada N, Villaroel M, María A. Pre-slaugther cattle welfare indicators for use in comercial abattoirs with voluntary monitoring systems: a systematic review. *Meat Sci.* (2017) 133:34–48. doi: 10.1016/j.meatsci.2017.12.004
28. Alam NG, Gregory MA, Jabbar MS, Uddin AS, Kibria A, Silva-Fletcher A. Skin injuries identified in cattle and water buffaloes at livestock markets in Bangladesh. *Vet Rec.* (2010) 167:415–9. doi: 10.1136/vr.c3301
29. Cobo CG. *Evaluación de la interacción hombre-animal como indicador de bienestar animal durante el presacrificio bovino* (Ph.D. thesis). Universidad de Caldas, Manzalez, Colombia (2013).
30. Herrán L, Romero M, Herrán L. Interacción humano-animal y prácticas de manejo bovino en subastas colombianas. *Rev Inv Vet Perú.* (2017) 28:571–85. doi: 10.15381/rivep.v28i3.13360
31. Van de Water G, Verjans F, Geers R. The effect of short distance transport under commercial. *Livest Production Sci.* (2003) 82:171–9. doi: 10.1016/S0301-6226(03)00010-1
32. Minka N, Ayo J. Effects of loading behaviour and road transport stress on traumatic injuries in cattle transported by road during the hot-dry season. *Livest Sci.* (2007) 107:91–5. doi: 10.1016/j.livsci.2006.10.013
33. Bourguet C, Deiss V, Cohen C, Terlouw E. Behavioural and physicological reaction of cattle in a commercial abattoir: Relationships with organizational aspects of the abattoir and animal characteristics. *Meat Sci.* (2011) 88:158–68. doi: 10.1016/j.meatsci.2010.12.017
34. Hultgren J, Wiberg S, Berg C, Cvek K, Lunner C. Cattle behaviours and stockperson actions related to impaired animal welfare at Swedish slaughter plants. *Appl Anim Beh Sci.* (2014) 152:23–37. doi: 10.1016/j.applanim.2013.12.005
35. Miranda-de la Lama G, Leyva I, Barrera-Serrano A, Pérez-Linares C, Sánchez-López E, María G, et al. Assesment of cattle welfare at a comercial slaughter plant in the northwest of México. *Tropical Anim Health Prod.* (2012) 44:497–504. doi: 10.1007/s11250-011-9925-y
36. Dahl-Pedersen K, Herskin M, Houe H, Thomsen P. A descriptive study of the clinical condition of cull dairy cows before transport to slaughter. *Livest Sci.* (2018) 218:108–13. doi: 10.1016/j.livsci.2018.11.001
37. Hernández-Gotelli C, Tadich N, Sepúlveda P. *Causas de eliminación de vacas lecheras en tres rebaños de la Región de los Ríos, Chile* (Ph.D. dissertation). XXIV Congreso de la Asociación Latinoamericana de Producción Animal, Puerto Varas (2015).
38. Karrar MH, Osman KhM, Sulieman MS. Culling in dairy cattle farms of Khartoum, Sudan. *Online J Anim Feed. Res.* (2017) 7:1–8. Available online at: https://www.researchgate.net/publication/313108785_CULLING_IN_DAIRY_CATTLE_FARMS_OF_KHARTOUM_SUDAN
39. Leach K, Knierim U, Whay HR. Condition scoring for dairy and beef cattle and veal calves. In: Forkman B, Keeling L, editor. *Welfare quality report n° 11 - Assessment of Animal Welfare Measures for Dairy Cattle, Beef Bulls and Veal Calves*. Cardiff: Cardiff University (2009). p. 1–6.
40. Moorman A, Duffield T, Godkin M, Kelton D, Rau J, Haley D. Associations between the general condition of culled dairy cows and selling price at Ontario auction markets. *J. Dairy Sci.* (2018) 101:10580–8. doi: 10.3168/jds.2018-14519
41. Grandin T. Behavioural principles of handling cattle and other grazing animals under extensive conditions. In: Grandin T, editor. *Livestock Handling and Transport*. Wallingford: CABI (2014). p. 39–64.

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.

frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 07 August 2020
doi: 10.3389/fvets.2020.00445

Animal Welfare Assessment Protocol for Does, Bucks, and Kit Rabbits Reared for Production

Antoni Dalmau, Xenia Moles and Joaquim Pallisera*

Animal Welfare Program, Institute of Agrifood Research and Technology, Girona, Spain

Rabbits are the third species in terms of number of animals reared for meat production in the world. However, in comparison to other species, very few studies have focused on their welfare. The aim of the present study is to implement an animal welfare-assessment protocol developed through a multidimensional approach and containing a number of animal-based measures for bucks, does, and kit rabbits. Thirty Spanish farms with conventional cages in the first year of integration in an animal welfare certification scheme were visited during 2019 and audited by the same auditor. The protocol is divided into four principles and 11 criteria. The Good Feeding principle includes eight parameters (one animal-based), Good Housing includes 15 parameters (six animal-based), Good Health includes 26 parameters (16 animal-based), and Appropriate Behavior contains nine parameters (four animal-based). In general, the main problems found were the absence of platforms, low space allowance and low height of the cage, inappropriate systems for performing emergency killing, insufficient protection of does from other adjacent does when housed individually, and absence of enrichment material. To a minor degree, but also found in an important number of farms, was a lack of temperature data records, high replacement rates of does, and lack of mortality-rate data records. However, in general, most of the farms obtained a good overall score, the maximum found being 73 out of 100 points. Nevertheless, none of the farms reached an excellent score, and four farms were scored below the 55 points required in the animal welfare certification scheme. The Good Feeding principle obtained the highest score, reaching excellent in all farms, and Appropriate Behavior the lowest one, with values ranging from 21 to 41 points out of 100. The results probably show how, for years, rabbit producers have been very focused on feeding needs and very little attention has been paid to behavioral needs.

Keywords: animal-based parameters, animal welfare, assessment protocol, behavior, health, housing, injuries, rabbits

OPEN ACCESS

Edited by:
Edward Narayan,
The University of Queensland, Australia

Reviewed by:
Elizabeth Anne Mcbride,
University of Southampton, United Kingdom
Alan G. McElligott,
University of Roehampton London, United Kingdom

****Correspondence:***
Antoni Dalmau
antoni.dalmau@irta.es

Specialty section:
This article was submitted to Animal Behavior and Welfare, a section of the journal Frontiers in Veterinary Science

Received: *24 April 2020*
Accepted: *19 June 2020*
Published: *07 August 2020*

Citation:
Dalmau A, Moles X and Pallisera J (2020) Animal Welfare Assessment Protocol for Does, Bucks, and Kit Rabbits Reared for Production. Front. Vet. Sci. 7:445. doi: 10.3389/fvets.2020.00445

INTRODUCTION

Domestic rabbits *(Oryctolagus cuniculus)* bred for meat production represent <1% of the meat produced in the world. However, although in terms of kg, rabbits are negligible in comparison to other species, animal welfare considers individuals and not tons of meat. According to the FAOSTAT, in 2018, the number of slaughtered animals in the world for meat production was 68.785 billion chickens, 1.484 billion pigs, 922 million rabbits, 656 million turkeys, 573 million sheep, 479 million goats, 302 million cattle, and 5 million horses (1). Therefore, rabbits are the third species in terms of number of animals reared for meat production in the world. As happens with other species, the production of these animals has been increasing in the last few years. From 2008 to 2018, the number of rabbits reared in the world for production increased 9.8% (1).

According to Broom (2), the welfare of an animal is its state as regards its attempts to cope with its environment, and stress response plays a key role in these attempts. In fact, stress has been defined as a state that occurs when an animal is required to make abnormal or extreme adjustments, in either its physiology or behavior, in order to cope with adverse aspects of its environment and management (3). Stress affects the animals in different ways, such as reduced feed intake (4, 5), increased disease susceptibility (6), reproductive efficiency (7, 8), or changing the behavior of animals (9, 10).

Probably due to the low economic impact of the production of rabbits worldwide or due to the geographical distribution of this production system, rabbits are by far the meat production species least studied in terms of animal welfare, particularly when compared to pigs, chickens, turkeys, cattle, sheep, and goats (11). For instance, the European Union (EU) funded from 2004 to 2009 one of the most ambitious projects on animal welfare ever developed, the Welfare Quality project. One of the aims of this project was to develop protocols to assess animal welfare in an objective, science-based, and practically applicable way, focusing the assessment on animal-based parameters (12). However, this project was focused only on cattle, pigs, and chickens. After this project, and following the principles stated by the Welfare Quality project, the EU funded a second one, named AWIN, which covered the omitted species in the previous one: turkeys, sheep, goats, and horses. Yet, rabbits were never considered in either of the two European projects. Nonetheless, rabbit producers are subjected to the same challenges as other producers: a higher demand from consumers for animal-friendly production systems and a greater production efficiency to increase marginal benefits. In both cases, a better knowledge of animal welfare and tools for their assessment plays a central role. These tools, or animal welfare-assessment protocols, can be used by farmers to identify critical points in the farm for investments, to compare their own results with those from other producers to perform self-assessments, and to create communication channels with the consumer to give an added value to the farms with better conditions.

De Jong et al. (13) presented a first step in the development of an animal welfare-assessment protocol for commercially housed rabbits. This consisted of describing possible parameters for the different criteria and principles as used in Welfare Quality project. This was done by combining the little information existing in the literature and the opinion of experts from different countries. This basis was tested in Spain to build a possible animal welfare protocol for rabbits and, from 2016 to 2018, it was tested in different rabbit farms for meat and fur production. According to the results obtained, some extra parameters were added, and others changed. Then the thresholds for the different measures and a score system were developed. The objective of the present study was to present the protocol based on the Welfare Quality approach for discussion after its implementation in 30 farms assessed in Spain that were interested in achieving certification on animal welfare for does, bucks, and kit rabbits.

METHODS

Thirty Spanish rabbit farms with conventional cages were visited from March to December 2019 and assessed by means of an animal welfare protocol for does, bucks, and kit rabbits in a single visit. These farms were audited within an animal welfare certification system (Welfair^{TM}; https://www.animalwelfair.com/) and assessed in all cases by the same auditor, who was trained according to the training procedure established in Welfare Quality (12). The system was presented to different national congresses and meetings and farmers volunteered to be audited as part of the study. In all cases, they had the opportunity to check and test the protocol for doing a self-assessment in their farms before the audit, and they were aware that 55 points out of 100 was the cut off for these protocols. However, it is unknown how many of them carried out this self-assessment, whether they used a part or all the measures of the protocol for this purpose, or if they just did an overview to the protocol before confirming their interest in being audited. In any case, the results presented in this study were collected during the first assessment done by an external auditor in all 30 farms assessed. Cages to be assessed were selected randomly throughout the buildings to be representative of the overall picture of the farm. Although the protocols can be used for bucks and does, only one of the assessed farms had males for reproduction, and in the rest, females were artificially inseminated. The Welfare Quality schema provides four different principles to assess animal welfare, and these are divided into 12 criteria (**Table 1**). This is the structure used in the present study. However, as in the present protocol it was not possible to identify any good parameter to assess the positive emotional state criterion, the protocol includes only 11 criteria. Globally, the Good Feeding principle includes eight parameters (one animal-based); Good Housing includes 15 parameters (six

TABLE 1 | Principles and criteria defined in the European Welfare Quality project to assess animal welfare (12).

Principles	Criteria
Good Feeding	Absence of prolonged hunger
	Absence of prolonged thirst
Good Housing	Comfort around resting
	Thermal comfort
	Ease of movement
Good Health	Absence of injuries
	Absence of diseases
	Absence of pain induced by management
Appropriate Behavior	Social behavior
	Other behaviors
	Human-animal relationship
	Positive emotional state

animal-based); Good Health includes 26 parameters (16 animal-based); and Appropriate Behavior contains nine parameters (four animal-based).

Good Feeding

The Good Feeding principle is assessed by means of the combination of two criteria: absence of prolonged hunger (65% of the total score) and absence of prolonged thirst (35% of the total score; **Table 2**). In relation to the first criterion, the protocol states that, for the body-condition parameter, a total of 24 bucks and 34 does are assessed (when possible, 17 around mounting or insemination and 17 just before weaning). If there are no males in the farm, a total of 50 does are assessed (when possible, 25 around insemination and 25 just before weaning). Presence of bad body condition (thinness) is assessed visually, considering an animal to be lean when hips and backbone are very prominent. The cleanliness of the feeders, number of drinking points per doe/buck, functioning of the drinkers, and cleanliness of the drinkers parameters are assessed in 24 cages for bucks and 51 for does (when possible, 17 in the first week post-partum, 17 around insemination, and 17 post-weaning). If the farm does not contain bucks, it is assessed in 75 cages for does (when possible, 25 of each type). A feeder is considered dirty when it contains corrupted food, compacted dry food, and mold. A bad functioning of the drinkers is considered when there is an insufficient flow or if they are dripping. Access to milk in kits is assessed by asking the farmer and, if possible, checking it during the visit, if all of the kits during the first 7 days of life are checked after the visit of the doe to the nest to ensure that all of them have had milk. Access to food by kits and height of drinkers are only assessed if kits of at least 21 days of age are present (**Table 2**). If they do not have access to solid food, 20 points are subtracted from the whole absence of prolonged hunger criterion. If the drinker for these kits is not <22 cm from the floor, 20 points are subtracted from the absence of prolonged thirst criterion.

Good Housing

The Good Housing principle is assessed by means of the combination of three criteria: comfort around resting (40%), thermal comfort (25%), and ease of movement (35% of the total score; **Table 3**). The wet animals, dirty animals, presence of resting mats, height of the cage/pen, and stocking density parameters are assessed in 24 cages for bucks and 51 for does (when possible, 17 in the first week post-partum, 17 around insemination, and 17 post-weaning). If the farm does not contain bucks, it is assessed in 75 cages for does (when possible, 25 of each type). A wet animal is considered when any part of the fur is wet. For dirtiness, two categories are considered. The animal is scored as moderately dirty when from 10 to 30% of the body is dirty, and severely dirty when more than 30% of the body is dirty. The stocking density is assessed in cm^2 of free space per animal. The parameters of free movement, panting, and shivering are assessed in a total of 10 bucks and 40 does not assessed for other parameters (if there are not bucks in the farm, 50 does are assessed). Each animal is assessed during a time of 2 min. Free movement is defined as the capacity of the animal for performing hopping, jumping, and turning. An animal is considered as panting when it is breathing with short

TABLE 2 | Parameters used to assess the criteria of absence of prolonged hunger and absence of prolonged thirst.

Criterion	Parameter	Weight (%)	Categories	Definition of categories	Score
Absence of prolonged hunger	Body condition	70	Excellent	0% of lean animals	100
			Acceptable	Up to 3% of lean animals	70
	Cleanliness of feeders	15	Excellent	100% of clean feeders	100
			Acceptable	At least 97% of clean feeders	45
	Access to milk in kits	15	Excellent	All kits checked after doe's visit the first week	100
			Not acceptable	Any kit not checked after doe's visit	0
	Access to food in kits	0	Acceptable	Solid food for kits of 21 days or older	0
		−20	Not acceptable	No solid food for any kit of 21 days or older	−20
Absence of prolonged thirst	Drinking points per doe/buck	45	Excellent	More than 1 drinking point per doe/buck	100
			Acceptable	A ratio of 1 drinking point for 1 doe/buck of fresh water tested and working	65
	Functioning of drinkers	35	Excellent	100% with a good water flow	100
			Acceptable	At least 97% with a good water flow	55
	Cleanliness of drinkers	25	Excellent	100% of clean drinkers	100
			Acceptable	At least 97% of clean drinkers	50
	Height of drinkers	0	Excellent	Drinker for kits at 13 cm from the floor or less	0
		−20	Acceptable	Drinker for kits at >13 and <23 cm from floor	−10
			Not Acceptable	Any other situation for kits older than 21 days	−20

The weight means which percentage of the score of the total criterion is represented by each parameter. Each parameter can be assessed according to different categories (excellent, acceptable, and not acceptable) and are scored accordingly.

TABLE 3 | Parameters used to assess the criteria of comfort around resting, thermal comfort, and ease of movement.

Criterion	Parameter	Weight (%)	Categories	Definition of categories	Score
Comfort around resting	Fully stretched animals	15	Excellent	At least 20% of the animals fully stretched	100
			Acceptable	At least 10% of the animals fully stretched	45
	Wet animals	20	Excellent	<5% of wet animals	100
			Acceptable	<10% of wet animals	60
	Dirty animals	20	Excellent	Up to 2% moderately and 0% severely dirty	100
			Acceptable	Up to 4% moderately and 2% severely dirty	60
	Dust	15	Excellent	No dust presence	100
			Acceptable	Minimal dust present	70
	Presence of Resting mat	30	Excellent	100% of does and bucks with resting mat	100
			Acceptable	At least 50% of does and bucks with resting mat	50
	Presence of an elevated platform	0	Excellent	100% of the cages with a platform	0
		−20	Acceptable	At least 50% of the cages with a platform	−10
			Not acceptable	Any other situation	−20
	Light quality	0	Acceptable	8 h of light and 8 h of darkness and enough light to check animals	0
		−20	Not acceptable	Any other situation	−20
	Quality of littered floor	0	Excellent	Clean and dry litter in all nests	0
		−20	Acceptable	No clean or dry litter in up to 2 nests	−10
			Not acceptable	Any other situation	−20
Thermal comfort	Temperature	100	Excellent	Last 3 months with range of 1–28°C	100
			Acceptable	Up to 2 days out of this range	50
	Burning hair	0	Acceptable	During burning hair not >28°C	0
		−20	Not acceptable	No data or >28°C	−20
	Panting	0	Excellent	0% of animals panting	0
		−100	Acceptable	Up to 4% of animals panting	−50
			Not acceptable	More than 4% of animals panting	−100
	Shivering	0	Excellent	0% of animals shivering	0
		−100	Acceptable	Up to 4% of animals shivering	−50
			Not acceptable	More than 4% of animals shivering	−100
Ease of movement	Free movement	30	Excellent	100% of the animals with free movement	100
			Acceptable	At least 97% of the animals	65
	Height of the cage	30	Excellent	38 cm at least in 90% of the cages	100
			Acceptable	32 cm at least in 90% of the cages	50
	Stocking density	40	Excellent	At least 3,500 cm^2 per doe/buck in 90% of cages	100
			Acceptable	At least 2,500 cm^2 per doe/buck in 90% of cages	60

The weight means which percentage of the score of the total criterion is represented by each parameter. Some parameters, such as light quality, are only considered in the score when the acceptable value is not achieved, subtracting points from the overall score. Each parameter can be assessed according to different categories and are scored accordingly.

and quick breaths and with the mouth open. The fully stretched lying animals parameter is assessed in all animals assessed for dirtiness and free movement, a total of 34 bucks and 91 does, or 125 does if there are no bucks in the farm. The quality littered floor parameter is assessed only in cages with does and kits in the first week after kindling (from 17 to 25 cages) and refers to the material present in the nest. The presence of an elevated platform parameter is assessed only in cages with kits older than 21 days (from 17 to 25 cages). The dust, light quality, environmental temperatures, and burning hair parameters are assessed globally in the facilities of each farm. The dust parameter is assessed by means of a black surface of ∼10 * 15 cm (DINA 4-folds in four pieces) and left during the assessment at the center of the building housing the rabbits at the same height as their heads. At the end of the visit, the level of dust accumulated is assessed considering three possibilities: no evidence of dust, minimal evidence of dust (a thin covering of dust), and a lot of dust (possible to write on the paper with a finger, or paper not visible). The quality of light is considered correct when it is possible to check all of the animals and if at least 8 h of light and darkness are provided. When it is considered unacceptable, 20 points are subtracted from the comfort around resting criterion. The temperature parameter is assessed according to the temperature data record in the farm. If there are no data, 0 points are given. If there are data, excellent is given when the temperatures, maximum and minimum, range from 1 to 28°C, respectively. Acceptable is considered when, up to 2 days in the last 3 months, the temperature registered is out of this range and, finally, unacceptable in any other case. The burning hair parameter is related to the burning of molted hair accumulated in the cages for improving environmental conditions. In this case, if the temperature is outside of the proposed ranges (1–28°C) during this practice or not registered, 20 points are subtracted from the thermal comfort criterion.

Good Health

The Good Health principle is assessed by means of the combination of three criteria: absence of injuries (40%), absence of diseases (40%), and absence of pain induced by management (20%, **Table 4**). The parameters of wounds on the body, wounds on the ears, fallen ears, pododermatitis, gait score, nasal discharge, ocular discharge, dermatophytosis/dermatitis/abscesses, neck torsions, enteropathy, diarrhea, mange, risk of injuries, and cleanliness of the housing system are assessed in 24 cages for bucks and 51 for does (when possible, 17 in the first week post-partum, 17 around insemination, and 17 post-weaning). If the farm does not contain bucks, it is assessed in 75 cages for does (when possible, 25 of each type). For wounds on the body, a lesion is considered a fresh scratch or open lesion equal or bigger than 2 cm in any part of the animal and not healed. Any animal with these lesions is assessed as moderately injured. Nevertheless, in case of a lesion of equal or more than 5 cm, the animal is assessed as severely injured. For wounds on the ears, no distinction for size is made, and only lesions of equal or more than 2 cm are considered. However, in this case, old lesions are distinguished from fresh lesions. Fallen ears are considered just as the absence or presence of the problem, and only the worse of the two ears is considered. This parameter is not considered for certain breeds with fallen ears, such as beliers. Pododermatitis considers three cases: no problem, when the feet are fine; moderate problems, if there is no hair, with a callus formed, and the area affected being longer than 2 cm; and severe problems, if there is an open lesion. Gait score considers three possibilities: no problems, if the animal does not have any difficulty in moving; moderate problem, if the animal has any difficulty in moving; and severe problem, if the animal has several difficulties (no use of one leg or minimum weight bearing). Nasal and ocular discharge is only considered as the presence or absence of the problem. Signs of conjunctivitis are considered as the presence of ocular discharge. Any sign of skin inflammation is considered, as the presence or absence, in the parameter of dermatophytosis/dermatitis/abscesses. For neck torsion, three conditions are considered: absence, when the neck is perfect; moderate problem, when the animal has a neck torsion but is able to eat and drink with no difficulties; and severe problem, when the neck torsion makes access to food and water difficult for the animal. Enteropathy is assessed by palpation of the abdomen and is considered present when this is hard. Diarrhea is assessed as the presence of liquid feces around the anus of the animal. Mange is assessed as its presence or absence. Risk of injuries is assessing the risk for the animals to be injured by bad maintenance of the cages or other elements in their surroundings. This parameter subtracts up to 30 points from the absence of injuries criterion if more than one cage with a problem is found. The cleanliness of the housing system parameter has three possibilities: the cage is clean; the cage is partly dirty, when only a part of the cage is affected (including a lot of presence of hair, compacted dry food, and mold); and a dirty cage, when the entire cage is very dirty. This parameter subtracts up to 20 points from the absence of diseases criterion. The coughing and sneezing parameters are assessed in a total of 10 bucks and 40 does not assessed for other parameters (if there are no bucks in the farm, 50 does are assessed). Each animal is assessed during a time of 2 min, and the presence or absence of coughing or sneezing during this period is considered. Hairless areas are assessed only in bucks (in a total of 24 animals) and considered present when there is an area of equal or more than 2 cm without hair. Mortality and culling rates are assessed according to the records of the farm in the last 3 months. Mortality considers only adult does and bucks deaths in the farm and not culled by the farmer. The parameters of replacement, time between parturitions, age of weaning, mutilations used for identification, and procedure for emergency killing are asked of the farmer and, when possible, assessed during the visit. The presence of flies parameter is assessed by observing the facilities where animals are housed, and three possibilities are considered: no problem, when no presence of flies or fly eggs is observed overall in the farm; moderate problems, when the presence of flies or fly eggs is observed (which indicates a problem with flies only in the past [eggs] or only in the present [flies]); and severe problem, when the presence of flies and fly eggs is observed (which indicates a chronic problem of flies in the farm).

Appropriate Behavior

The Appropriate Behavior principle is assessed by means of the combination of three criteria: social behavior (35%), other behaviors (35%), and human-animal relationship (30%, **Table 5**). The parameters of negative social behavior, abnormal behaviors, and human approach test are assessed in a total of 10 bucks and 40 does not assessed for other parameters (if there are no bucks in the farm, 50 does are assessed). Negative social behavior is considered as any such event in which a doe or a buck is biting another one, including kits in the same cage or any animal in other adjoining cages for a time of 2 min. Abnormal behavior consists of animals scratching or biting the cage or performing repetitive behaviors without an apparent objective (stereotypies) and it is assessed for a time of 2 min per cage. The human approach test is performed after the 2 min dedicated to assess the other behavioral measures. For 30 s, the assessor would be in front of the cage of the animal touching the frontal area of the cage with a short stick (no more than 10 cm long, a new stick is used for each animal). Three possibilities are considered: confident, if the animal touches or sniffs the stick; interested, if the animal shows some interest in the stick and approaches to at least 10 cm from the stick; not interested or fearful, any other situation. The isolated animals and presence of enrichment material parameters are assessed in a total of 34 bucks and 91 does (or 125 does is

TABLE 4 | Parameters used to assess the criteria of absence of injuries, absence of diseases, and pain induced by management procedures.

Criterion	Parameter	Weight (%)	Categories	Definition of categories	Score
Absence of injuries	Wounds on the body	25	Excellent	Up to 2% moderate and 0% severe	100
			Acceptable	Up to 4% moderate and 2% severe	60
	Wounds on the ears	15	Excellent	Up to 2% old and 0% fresh lesions	100
			Acceptable	Up to 4% old and 2% fresh lesions	45
	Fallen ears	10	Excellent	Up to 2% with fallen ears	100
			Acceptable	Up to 4% with fallen ears	70
	Pododermatitis	30	Excellent	Up to 50% moderate and 5% severe	100
			Acceptable	Up to 65% moderate and 8% severe	50
	Gait score	20	Excellent	Up to 2% moderate and 0% severe	100
			Acceptable	Up to 4% moderate and 2% severe	50
	Hairless areas	0	Excellent	0% of animals affected	0
		−10	Acceptable	Up to 13% of animals affected	−5
			Not acceptable	More than 13% of animals affected	−10
	Risk of injuries	0	Acceptable	No cages with risk of injuries	0
		−30	Not accept	1 cage with risk of injuries	−15
			Not accept	More than 1 cage with risk of injuries	−30
Absence of diseases	Mortality	10	Excellent	Up to 3% in the last 3 months	100
			Acceptable	Up to 5% in the last 3 months	70
	Culling	5	Excellent	At least equal to or higher than mortalities	100
			Acceptable	At least 50% of mortality rates	40
	Replacement	5	Excellent	<80% of females per year	100
			Acceptable	<110% of females per year	40
	Time between parturitions	5	Excellent	At least 49 days between parturitions	100
			Acceptable	At least 42 days between parturitions	40
	Coughing	10	Excellent	0% of animals coughing during 2 min	100
			Acceptable	Up to 3% of animals coughing	70
	Sneezing	10	Excellent	0% of animals sneezing during 2 min	100
			Acceptable	Up to 3% of animals sneezing	70
	Nasal discharge	7	Excellent	Up to 2% of animals affected	100
			Acceptable	Up to 4% of animals affected	40
	Ocular discharge	8	Excellent	Up to 2% of animals affected	100
			Acceptable	Up to 4% of animals affected	50
	Dermatophytosis, dermatitis, abscesses	10	Excellent	0% of dermatophytosis and up to 4% of animals with dermatitis or abscesses	100
			Acceptable	0% of animals with dermatophytosis and up to 10% of animals with dermatitis or abscesses	70
	Neck torsions	10	Excellent	Up to 2% moderate, 0% severe	100
			Acceptable	Up to 4% moderate, 2% severe	70
	Enteropathy	10	Excellent	0% of animals affected	100
			Acceptable	Up to 2% of animals affected	70
	Diarrhea	10	Excellent	0% of animals affected	100
			Acceptable	Up to 4% of animals affected	70
	Mange	0	Acceptable	0% of animals affected	0
		−40	Not acceptable	At least one animal affected	−40
	Cleanliness of facilities	0	Acceptable	Up to 2 cages partly dirty	0
		−20	Not acceptable	Up to 5 cages partly dirty and up to 2 cages dirty	−10
			Not acceptable	Any other case	−20
	Age of weaning	0	Acceptable	35 days old or older	0
		−20	Not acceptable	Before to 35 days old	−20

(Continued)

TABLE 4 | Continued

Criterion	Parameter	Weight (%)	Categories	Definition of categories	Score
	Flies presence	0	Acceptable	No flies neither eggs of flies present	0
		−20	Not acceptable	Flies or eggs of flies present	−10
			Not acceptable	Flies and eggs of flies present	−20
Absence of pain induced by management	Killing methods	100	Excellent	Penetrative captive bolt with pithing, penetrative captive bolt with bleeding or penetrative captive bolt with neck dislocation Electronarcosis with neck dislocation or electronarcosis with bleeding Lethal injection	100
			Not acceptable	None of the previous systems	0
	Mutilations for identification	0	Excellent	No mutilations performed	0
		−20	Not acceptable	Mutilations for identification in any part of the body	−20

The weight means which percentage of the score of the total criterion is represented by each parameter. Some parameters, such as risk of injuries, are only considered in the score when the acceptable value is not achieved, subtracting points from the overall score. Each parameter can be assessed according to different categories and are scored accordingly. In some cases, the category not acceptable can have more than one score.

there are no bucks in the farm). Isolated animals are positively considered when adult animals, due to the distance between cages or to the provision of solid separations, are protected from physical contact from other adults. In addition, it is negatively considered if any animal is visually isolated from other animals. The presence of enrichment material, different to metal cans, such as cubes of dried hay or wood sticks, is considered as being present or absent. The availability of nesting material parameter will be assessed, when possible, in 10 does 24 h prior to the date expected for kindling by checking for the provision of enough dry and clean nesting material. The parameters of training of personnel, time of having access to the nest, and touching kits every day are assessed by asking the farmer, although the last two parameters can be checked as well during the audit. Training of personnel considers three levels: all personnel in the farm in contact with the animals are trained in animal welfare; at least one person is trained in animal welfare; and none of the persons is trained in animal welfare. Certificates for any training must be shown. Time of accessing the nest assesses the regularity in giving the doe access every day to the nest at the same hour and considers three categories: no problem, when <1 h of difference is found among different days; moderate problems, when a difference among days of more than 1 h but <2 h is found; and severe problems, when the difference is of more than 2 h. Touching the kits every day assesses whether the manager is touching the kits for the first week of age at least once a day.

Overall Assessment

The Welfare Quality provides a final score for a farm as a result of the combination of the scores of the different principles. When the final score is between 0 and 19 points, the farm is considered not classified. When the final score is between 20 and 54 points, the farm is considered acceptable. When the final score is between 55 and 79 points, the farm is considered enhanced; and, finally, a farm from 80 to 100 points is considered excellent. In the present protocol, the same system was adopted, but changing the category not classified to not acceptable. For an overall assessment of a farm, the four principles of the rabbit protocol have different weights because of the number or importance of the measures included in the specific principles (8, 15, 26, and 9 parameters for principle, respectively). Consequently, to obtain the final score of a farm, 15% depends on the Good Feeding principle, 30% on the Good Housing principle, 35% on the Good Health principle, and 20% on the Appropriate Behavior principle. The overall score can range from 0 to 100 points.

RESULTS

Good Feeding

The Good Feeding principle is assessed by means of the combination of two criteria: absence of prolonged hunger and absence of prolonged thirst. Ninety-seven percent of the farms ($n = 29$) obtained 100 points for the absence of prolonged hunger criterion. That means that all does or bucks were found with a good body condition, the feeder was clean, kits were checked once a day to ensure that milk was taken in during the first 7 days of life, and that solid food was provided at least at 21 days of age. In one farm 20% of the feeders were found dirty and was scored with a 0 for this parameter. Consequently, the absence of prolonged hunger criterion obtained 85 points for this farm, and the 30 farms were classified as excellent for the criterion.

In the case of the absence of prolonged thirst criterion, only one farm had more than one drinking point per adult animal, obtaining the 100 points in this parameter. All of the rest of the farms had at least one drinker tested and working properly per animal, so they obtained 65 points. No problems of insufficient flow or dripping were found in any drinker of the 30 farms, so all of them obtained 100 points. Only in one farm was a problem found of dirtiness in the drinkers, reaching 8% of the drinkers assessed (higher than the acceptable level of 3%), so 0 points were given for this parameter to this farm. In relation to the height of the drinkers for the kits after 21 days of age, in 50% of the cases ($n = 15$), the drinker was below 14 cm from the floor, obtaining an excellent.

TABLE 5 | Parameters used to assess the criteria of social behavior, other behaviors and human-animal relationship.

Criterion	Parameter	Weight (%)	Categories	Definition of categories	Score
Social behavior	Negative social behavior	100	Excellent	No animals biting other animals	100
			Not acceptable	One animal biting another animal	0
	Isolated animals	0	Excellent	No does/bucks visually isolated, but all of them physically isolated from other adults	0
		−100	Not acceptable	Up to 5% of does/bucks visually isolated and at least 50% physically isolated	−50
			Not acceptable	Any other situation	−100
Other behaviors	Abnormal behavior	60	Excellent	0% with abnormal behavior	100
			Acceptable	Up to 4% of animals affected	55
	Enrichment material	40	Excellent	100% with enrichment material	100
			Acceptable	At least 50% with enrichment material	60
	Nesting material	0	Excellent	Nesting material in 100% of the cases	0
		−20	Acceptable	Up to 1 doe without nesting material	−10
			Not acceptable	Any other case	−20
	Time of access to the nest	0	Excellent	<1 h of difference day to day	0
		−20	Acceptable	>1 and <2 h of difference day to day	−10
			Not acceptable	Any other case	−20
Human-animal relationship	Human approach test	70	Excellent	At least 20% touching the stick and 40% not touching the stick but interested	100
			Acceptable	At least 10% touching the stick and 20% not touching the stick but interested	50
	Training of personnel	30	Excellent	All personnel in contact with animals trained in animal welfare	100
			Acceptable	At least one person trained in animal welfare	50
	Touching the kits	0	Excellent	All kits in the nests are touched once a day	0
		−10	Not acceptable	Not all kits are touched	−10

The weight means which percentage of the score of the total criterion is represented by each parameter. Some parameters, such as isolated animals, are only considered in the score when the acceptable value is not achieved, subtracting points from the overall score. In some cases, the category Not acceptable can have more than one score.

In 37% of the cases ($n = 11$), the drinker was between 14 and 22 cm from the floor and 10 points were subtracted from the final score of the criterion; finally, in 13% ($n = 4$) the height was higher than 22 cm, the maximum being found 27 cm from the floor, and 20 points were subtracted from the whole criterion. Globally, 46% of the farms ($n = 14$) were classified as excellent for this criterion, with scores ranging from 85 to 90 points. The other 64% of the farms ($n = 16$) were classified as enhanced for the criterion, with scores ranging from 65 to 75 points.

When the score of the whole principle was considered, all farms obtained an excellent, with scores ranging from 85 to 97 points.

Good Housing

The Good Housing principle is assessed by means of the combination of three criteria: comfort around resting, thermal comfort, and ease of movement.

In relation to the comfort around resting criterion, 80% of the farms ($n = 24$) had at least 20% of the animals fully stretched, this percentage ranging from 20 to 65% of the animals. Twenty-seven percent had from 20 to 25% of animals stretched; 17% from 25 to 30%; 27% from 31 to 35%; and the remaining 9% more than 36% of animals fully stretched. In all of these cases, this parameter was scored with 100 points. Ten percent of the farms ($n = 3$) had more than 10% of fully stretched animals and <20%, obtaining 45 points for this parameter. Finally, another 10% of the farms had <10% of animals fully stretched, the minimum percentage observed being 3%. In these cases, the score for the parameter was 0 points. No wet or dirty animals were observed in any of the 30 farms assessed; consequently, 100 points were obtained in all cases for these two parameters. In 90% of the farms ($n = 27$), all of the assessed animals had resting mats in good conditions of maintenance and 100 points were given. In one farm, 63% of the animals had resting mats and 50 points were given. In other two farms <50% of the animals had a resting mat, and 0 points were given. Both dust and quality of littered floor obtained an excellent in all 30 farms assessed. Light quality was excellent in 87% of the farms ($n = 26$), but in the other four farms <8 h of light were provided and, consequently, 20 points were subtracted from the criterion. In addition, none of the 30 farms had an elevated platform for does with kits at 21 days of age or older, so 20 points were subtracted from the criterion for all the farms. Globally, 63% ($n = 19$) of the farms were classified as excellent (just 80 points in all cases); 27% ($n = 8$) of the farms were classified as enhanced (scores ranging from 65 to 72 points);

and 10% ($n = 3$) were classified as acceptable (ranging from 35 to 52 points).

In relation to the thermal comfort criterion, none of the animals assessed in any of the farms were found panting or shivering. However, in only 67% of the farms ($n = 20$) was there a data record of at least 3 months of daily maximum and minimum temperatures, so, as the values were within the range of 1–28°C, these farms obtained 100 points and were classified as excellent for the criterion. The rest 33% ($n = 10$) obtained 0 points and were classified as not acceptable for the criterion.

In relation to the ease of movement criterion, 80% of the farms ($n = 24$) had all of the animals with free movement and 100 points were obtained, while the rest 20% ($n = 6$) had <97% of the animals in these conditions, and 0 points were obtained. The height of the cages was of at least 38 cm in 30% of the farms ($n = 9$), the maximum being 40 cm of height (**Figure 1**). These farms were scored with 100 points. In 60% of the farms ($n = 18$), the height was exactly 32 cm, obtaining 50 points. Finally, in 10% of the farms ($n = 3$) the height was below 32 cm (ranging from 22 to 28 cm), and they were scored with a 0. In 50% of the farms ($n = 15$), at least 3,500 cm^2 per doe/buck were provided, ranging from 3,500 to 4,000 cm^2, and were scored with 100 points (**Figure 1**). In 37% of the farms ($n = 11$) at least 2,500 cm^2 per adult were given, ranging from 2,900 to 3,450 cm^2, and were scored with 60 points. Thirteen percent ($n = 4$) provided <2,500 cm^2 to the animals, ranging from 1,600 to 2,400 cm^2, and were scored with 0 points. Globally, 50% of the farms ($n = 15$) were classified as excellent for this criterion, with scores ranging from 85 to 100 points. Thirty percent of the farms ($n = 9$) were classified as enhanced, with scores ranging from 55 to 75 points. Thirteen percent of the farms ($n = 4$) were classified as acceptable, with scores ranging from 20 to 45 points. Finally, 7% ($n = 2$) were classified as not acceptable for ease of movement, with a score in both cases of 15 points.

When the score of the Good Housing principle was considered, 43% ($n = 13$) of the farms obtained an excellent (values ranging from 83 to 92 points). Forty-three percent ($n = 13$) of the farms obtained an enhanced (values ranging from 55 to 78 points). Finally, 13% ($n = 4$) of the farms obtained an acceptable (values ranging from 28 to 51 points).

Good Health

The Good Health principle is assessed by means of the combination of three criteria: absence of injuries, absence of diseases, and absence of pain induced by management.

In relation to the absence of injuries criterion, all farms assessed had 0% of animals with wounds on the ears, fallen ears, or gait score, and the only farm where bucks were present also had 0% of the animals with hairless areas. As a result, the score of excellent (100 points) was obtained for all of these farms in these parameters. In addition, 87% of the farms ($n = 26$) had an excellent for wounds on the body (100 points), two other farms had up to 4% of animals moderately injured (60 points), and the other two had 6 and 12% of animals moderately injured, respectively (0 points). None of the farms had severely injured animals. In the case of pododermatitis, the moderate cases ranged from 10 to 40% of the animals, the severe cases being what determined the score for the farm. In two farms, the percentage of severe cases of pododermatitis was 0% and in 12 other farms the percent was below 5%; consequently, 47% of the farms were scored with an excellent. Twenty-three percent of the farms

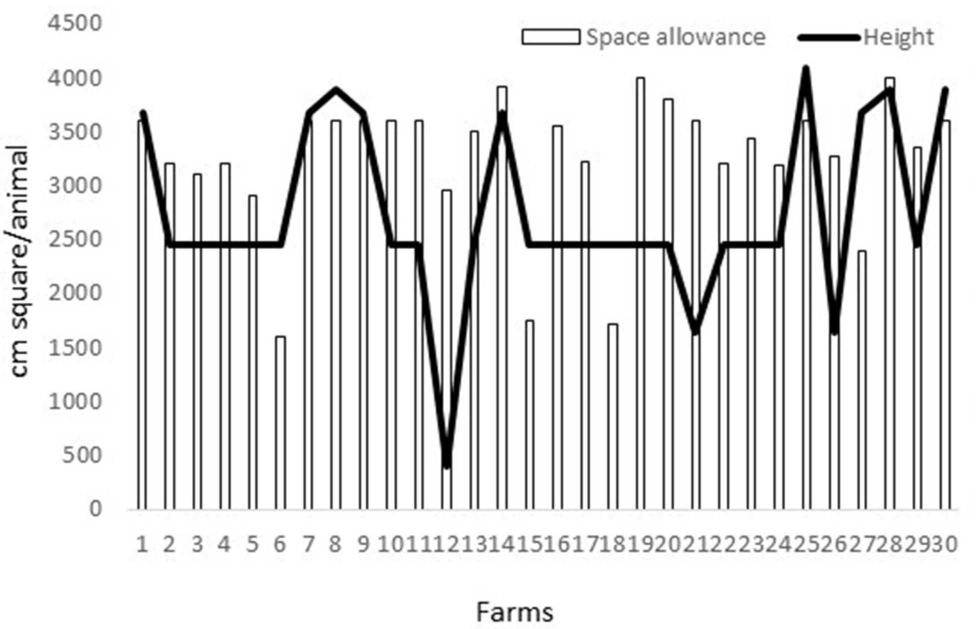


FIGURE 1 | Space allowance (cm^2/doe or buck; with bars) and height (cm; black line) found in the 30 Spanish farms of bucks, does, and kit rabbits assessed during 2019.

($n = 7$) were between 6 and 8% of severe cases of pododermatitis, being scored with 50 points. Finally, 30% of the farms ($n = 30$) had more than 8% of severe cases, the maximum found being 20% of animals affected. In these cases, the farms were scored with 0 points for this parameter. Seventy-three percent of the farms ($n = 22$) had no problems in any of their cages, regarding having a risk of injuries for the animals. However, for the rest of the farms, 27% ($n = 8$), at least two cages with a risk of injuries were found, so 30 points were subtracted from the whole criterion of absence of injuries. Globally, 33% of the farms ($n = 10$) obtained 100 points for this criterion, and 4 other farms obtained at least 85 points, so 47% of the farms were classified as an excellent for comfort around resting to absence of injuries. Forty percent of the farms ($n = 12$) were classified as enhanced (scores ranging from 55 to 70 points). Finally, 13% of the farms ($n = 4$) were classified as acceptable (scores ranging from 40 to 45 points).

In relation to the absence of diseases criterion, the parameters for coughing, sneezing, ocular discharge, neck torsion, and cleanliness of facilities obtained an excellent in all farms assessed, with no problems observed in any animal or cage. Seventeen percent of the farms ($n = 5$) obtained an excellent for mortality, with values ranging from 2.6 to 3%, and 33% of the farms ($n = 10$) obtained 70 points, ranging from 3.5 to 4.8% (**Figure 2**). Finally, 33% of the farms ($n = 10$) were between 5.2 and 25% of morality and another 17% ($n = 5$) had no data, in both cases being scored with a 0. Thirty percent of the farms ($n = 9$) had equal or higher percentages of animals culled as mortality rates. Thirty-seven percent of the farms ($n = 11$) had culling rates below 50% of mortality rates and another 33% ($n = 10$) had no data on culling rates, so 70% of the farms had a 0 for this parameter (**Figure 2**). Thirty percent of the farms ($n = 4$) had a replacement percentage of <80% per year, ranging from 35 to 70%, and were scored with 100 points. Another 17% of the farms ($n = 5$) had a replacement percentage of <110% per year, ranging from 90 to 108%, and scored with 40 points. Finally, 70% of the farms ($n = 21$) were scored with 0 points, ranging from 118 to 140% of replacement per year. Thirteen percent of the farms ($n = 4$) had at least 49 days between parturitions, ranging from 49 to 90 days, and scored with 100 points. All the rest of the farms had at least 42 days between parturitions and received 40 points. Eighty-three percent of the farms ($n = 25$) had up to 2% of animals with nasal discharge, being scored with 100 points. In the rest of the farms, 17% ($n = 5$) had more than 4% of animals with nasal discharge, ranging from 8 to 12% of animals affected, and were given 0 points. Ninety-seven percent of the farms ($n = 29$) did not have any problem with dermatophytosis, dermatitis, or abscesses, and they were scored with 100 points, but in one farm 8% of the animals were found with dermatitis, and the farm was given 70 points. Ninety-seven percent of the farms ($n = 29$) did not have any problem with enteropathy, and were scored with 100 points, but in one farm, 8% of the animals were affected, and the farm was scored with 0 points. Forty percent of the farms ($n = 12$) had 0% of animals with diarrhea and were scored with 100 points. Ten percent of the farms ($n = 3$) had up to 4% of the animals with diarrhea, being scored with 70 points. Finally, 50% of the farms ($n = 15$) had more than 4% of animals with diarrhea, ranging from 8 to 20%, and were scored with 0 points. Ninety-seven percent of the farms ($n = 29$) had no animals with mange, so an excellent was obtained for this parameter, but in one farm 45% of the animals were affected with mange and 40 points were subtracted from the whole criterion for this farm.

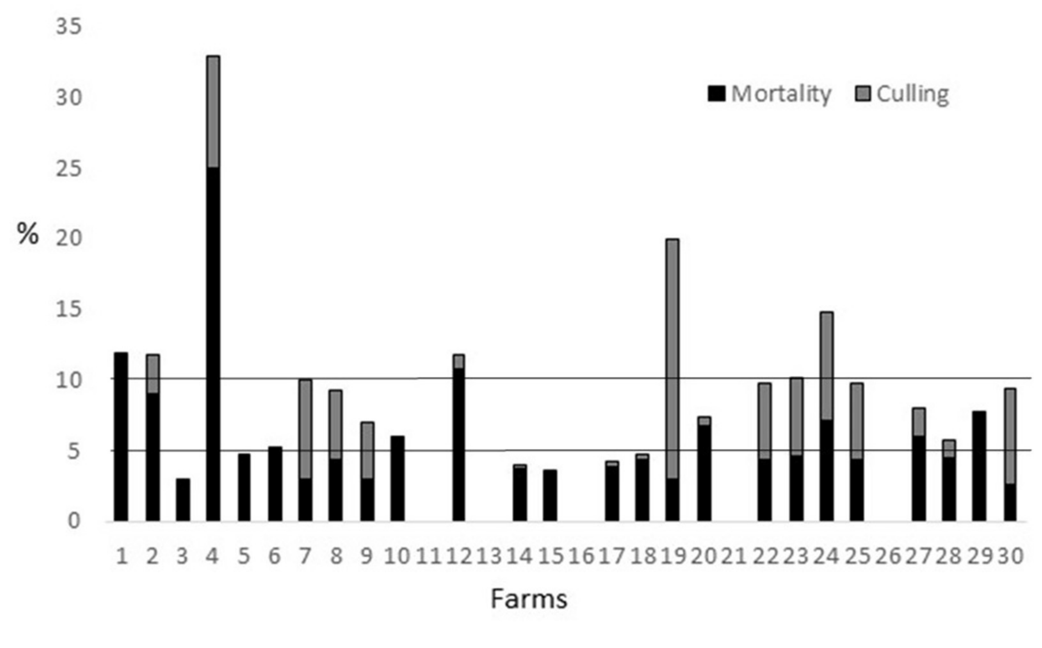


FIGURE 2 | Percentage of mortality (black part of the bar; %) and percentage of culling (gray part of the bar, %) registered during a period of 3 months in the 30 Spanish farms studied during 2019. The absence of a bar (black, gray, or both) means absence of data records for this farm.

Sixty-three percent of the farms ($n = 19$) did the weaning when kits were 35 days of age or older, ranging from 35 to 37 days, and scored with an excellent, while the rest 37% were weaned between 31 and 33 days old, and 20 points were taken from the whole criterion for these farms. Seventy-three percent of the farms ($n = 22$) had neither flies nor fly eggs, so they were scored with an excellent, while the rest, 27%, had flies and remains of eggs, so 20 points were subtracted from the whole criterion for these farms. In total, 27% of the farms ($n = 8$) were classified as excellent for the criterion (scores ranging from 80 to 95 points). Fifty percent of the farms ($n = 15$) were classified as enhanced for the criterion (scores ranging from 57 to 79 points). Finally, 23% of the farms ($n = 8$) were classified as acceptable for the absence of diseases criterion (scores ranging from 40 to 47 points).

In relation to the criterion of absence of pain induced by management, only two of the farms performed the emergency killing with any of the methods considered in the protocol, and were scored with 100 points. The farmers not using any of these systems were using cervical dislocation for adults or blunt force on the head for very young animals, and were scored with 0 points. In addition, in one of the farms the animals were mutilated in the ears for identification. The rest were using tattoos. Overall, 6% of the farms ($n = 2$) were classified with an excellent, and the rest with not acceptable for this criterion.

When the score of the whole principle was considered, only one farm had an excellent for the good health principle (scored with 86 points). Sixty percent of the farms ($n = 18$) were classified with an enhanced for this principle, ranging from 56 to 78 points. Finally, 37% of the farms ($n = 11$) were classified as acceptable for the Good Health principle, ranging from 40 to 54 points.

Appropriate Behavior

The Appropriate Behavior principle is assessed by means of the combination of three criteria: social behavior, other behaviors, and human-animal relationship.

In relation to the criterion of social behavior, no animals biting other animals were found in any of the farms assessed, so 100 points were obtained in all cases for this parameter. However, none of the animals were sufficiently protected from other adults due to the short distance between cages and the lack of a solid wall to protect the animals from being bitten, so 100 points to the whole criterion was subtracted in all cases. Consequently, all of the farms had 0 points and were classified as not acceptable for the social behavior criterion. However, no animals visually isolated from other animals were found in any farm.

In relation to the criterion of other behaviors, none of the animals in any of the farms assessed were found performing abnormal behaviors, so an excellent (100 points) was obtained for this parameter in all cases. On the other hand, none of the farms provided enrichment material to their animals, so 0 points were obtained for this parameter in all cases. In the farms where this could be assessed, does were provided with nesting material of good quality 24 h before kindling, so the excellent was always obtained in this parameter. In addition, the time of the day when the nest was open to the doe for suckling was every day at the same hour and with the same order in all farms assessed, so an excellent was obtained for this parameter. Globally, all farms were classified as enhanced for this criterion, with 60 points in total.

In relation to the human-animal relationship criterion and, more specifically, to the human approach test, the percentage of animals touching the stick was, in all cases, the reaction of the animal that defined the final score of the parameter due to the limited percentage of animals doing it. In fact, 20% of animals touching the stick was achieved only in one farm, and consequently scored with 100 points for this parameter. Thirty percent of the farms ($n = 9$) had at least 9% of the animals touching the stick, with percentages ranging from 10 to 18%. The rest of the farms, 67% ($n = 20$), were scored with 0 points, with percentages ranging from 0 to 8%. Thirty-three percent of the farms ($n = 10$) had all of the personnel in contact with the animals trained in specific courses on animal welfare, and they were scored with 100 points for this parameter. Another 10% of the farms ($n = 3$) did not have all the personnel, but at least one person, trained on animal welfare, and were scored with 50 points. The rest of the farms, 57% ($n = 17$) did not have any training for any of their personnel on animal welfare. In all farms, kits were touched once a day, so the excellent was obtained for this parameter. Globally, 7% of the farms ($n = 2$) were classified with an enhanced for this criterion (65 points). Forty-three percent of the farms ($n = 13$) were classified with an acceptable for this criterion (score ranging from 30 to 50 points), and 50% of the farms ($n = 15$) were classified as not acceptable for human-animal relationship (score ranging from 0 to 15 points).

When the score of the whole principle was considered, all of the farms were classified as acceptable for Appropriate Behavior, with values ranging from 21 to 41 points.

Overall Assessment

Considering a global score for all of the farms, where Good Feeding provides 15%, Good Housing 30%, Good Health 35%, and Appropriate Behavior 20% of the final score, none of the farms was classified as excellent (**Table 6**). Eighty-three percent of the farms ($n = 25$) were classified as enhanced, with scores ranging from 56 to 74 points. Finally, 17% of the farms ($n = 5$) were classified as acceptable, with scores ranging from 41 to 54 points.

DISCUSSION

The 30 farms assessed in the present study were all in the first year of integration in an animal welfare certification scheme, so the results could be biased in two ways. First, the producers know the protocol that will be used to assess their farms in advance and they enter the scheme voluntarily, so it is supposed that only those farmers really convinced of the capacities of their own farm to be approved would enter the scheme. Second, to be approved in this scheme a score of enhanced is needed in the global score, so only in cases of a very bad self-assessment of the farm is it expected to have global values below 55 points. Considering these two points, the objective of this discussion is to highlight the strengths and weaknesses of the farms assessed, the parameters used, and the protocol.

TABLE 6 | Score obtained (from 0 to 100 points) for the different criteria (C1: Absence of prolonged hunger; C2: Absence of prolonged thirst; C3: Comfort around resting; C4: Thermal comfort; C5: Ease of movement; C6: Absence of injuries; C7: Absence of diseases; C8: Pain induced by management; C9: Social behavior; C10: Other behaviors; C11: Human-animal relationship), for the different principles (P1: Good feeding; P2: Good housing; P3: Good health; P4: Appropriate behavior) and the overall score as the result of the combination of the scores of the 4 principles (TOTAL) by farm (from 1 to 30).

Farm	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	P1	P2	P3	P4	Total
1	100	75	35	0	100	55	82	0	0	60	0	91	49	55	21	52
2	100	65	80	100	75	60	67	0	0	60	0	88	83	51	21	60
3	100	85	65	0	75	55	75	0	0	60	35	95	52	52	32	54
4	100	65	80	100	75	90	64	0	0	60	35	88	83	62	32	66
5	100	65	80	100	75	100	67	0	0	60	0	88	83	67	21	66
6	100	75	80	100	15	100	40	0	0	60	30	91	62	56	30	58
7	100	65	80	100	100	100	95	0	0	60	35	88	92	78	32	74
8	100	75	80	100	100	45	72	0	0	60	35	91	92	47	32	64
9	100	90	65	0	100	100	92	0	0	60	0	97	61	77	21	64
10	100	75	80	0	90	100	85	0	0	60	35	91	64	74	32	65
11	100	85	80	0	90	55	57	100	0	60	50	95	64	65	36	64
12	100	75	80	100	40	60	47	0	0	60	0	91	71	43	21	54
13	100	85	80	0	90	100	57	0	0	60	35	95	64	63	32	62
14	100	85	80	100	100	70	72	0	0	60	0	95	92	57	21	66
15	100	75	72	100	20	70	47	0	0	60	65	91	61	47	41	56
16	100	75	60	0	90	40	60	0	0	60	0	91	56	40	21	48
17	100	75	80	100	45	70	74	0	0	60	0	91	73	58	21	60
18	100	75	80	100	15	85	47	0	0	60	30	91	62	53	30	57
19	100	85	80	0	90	85	90	0	0	60	0	95	64	70	21	62
20	100	75	72	100	85	60	47	0	0	60	65	91	84	43	41	62
21	100	65	35	0	40	70	40	0	0	60	0	86	28	44	21	41
22	100	85	60	100	100	70	79	0	0	60	30	88	84	60	30	65
23	100	85	80	100	75	100	91	0	0	60	30	95	83	76	30	72
24	100	85	80	100	75	70	80	0	0	60	0	95	83	60	21	64
25	100	65	60	100	100	70	79	0	0	60	30	88	84	60	30	65
26	100	85	80	0	55	100	47	0	0	60	50	95	51	59	36	57
27	100	85	80	100	60	100	65	100	0	60	15	95	78	86	26	73
28	100	85	80	100	100	85	62	0	0	60	15	95	92	59	26	67
29	100	85	52	100	75	100	80	0	0	60	0	95	72	72	21	65
30	85	85	60	100	100	40	71	0	0	60	30	85	84	44	30	59

Good Feeding

Within the Good Feeding principle, only one animal-based parameter is used: body condition score. According to the study of Bonanno et al. (14), the body condition score at artificial insemination (around 11 days post-partum) is a reliable indicator. However, later in the production cycle, when lactating does have simultaneous energy requirements for lactation and pregnancy, is as well a good moment for assessing body condition. For this reason, in the present protocol this parameter is assessed in animals around insemination and just before weaning. The way to assess this is based on Popescu et al. (15), where five categories were considered (emaciated, lean, ideal, fat, and obese), and simplified according to Welfare Quality (16), with just two categories, good (ideal) and lean (lean and emaciated). Popescu et al. (15) found more than 20% of the animals with problems of body condition in the two farms assessed, while in the present study 0% of animals affected in 30 farms was found. However, according to the authors, the low body-condition score found in their farms had, as the most probable cause, health-related reasons, and the present study had very good results for most of the parameters related to diseases. This, combined with the high replacement rates reported in some farms and the good hygienic conditions found, could explain these extremely good values in the body-condition score. Feeders should be positioned so that rabbits can easily access them while ensuring the feed is not contaminated by manure or urine. In 29 of 30 farms the cleanliness of feeders was excellent, so this seems to be under control in the farms assessed. According to NFACC (17), visually assessing kits for a milk spot in the first 5 days is a practical and effective way of confirming milk intake in kits. In the present study, all farmers carried out this practice for at least the first 7 days of life of the animals. In addition, all farmers were providing solid feed as well to the kits at 21 days of age. Kits gradually begin eating solid feed around 16–18 days of age, but usually their intake is in the form of milk until 25 days of age (17). Therefore, although the demand for providing solid feed could be advanced some days, it is not clear if this would really benefit the kit rabbits. Globally, this criterion is the one with the highest scores of the protocol. Although it could be argued that the absence of hunger and malnutrition is a basic principle for a professional rabbit producer, the low variability found could be caused by a low sensitivity of the parameters selected in detecting minor problems, so probably an effort is needed to include other parameters or to make the ones already included more sophisticated.

Rabbits have high water requirements and consume approximately twice as much water as feed, but there are no animal-based indicators available to measure prolonged thirst in a farm by visual inspection. Also, for other species, no animal-based parameters are currently in use for this criterion, and alternatively the number of drinking points and sometimes also the cleanliness and functioning of drinkers is assessed (16). Rabbits must have continuous access to safe, clean drinking water. In fact, it is important that nipples are clean, e.g., no hairs visible or mold, and working perfectly. In the farms assessed, in all cases except one, the drinkers were clean and all animals had at least one drinker working properly. In this criterion, the highest variability between farms was found in the height of the drinkers. The smallest weaned rabbits should also be able to drink, so the nipples should not be too high, an excellent height being when they are <14 cm from the floor, something found in 50% of the farms. Thus, although no animals could be considered thirsty during the assessment, this criterion (which could be called, as well, access to water) showed a higher variability between farms than did absence of prolonged hunger.

Good Housing

Facility design and maintenance significantly impacts rabbit health and welfare. Housing systems need to provide a comfortable environment for rabbits through appropriate space allowance. Floor space impacts a rabbit's ability to thermoregulate in high ambient temperatures [rabbits can cool themselves by stretching out; (17)]. In fact, rabbits' use of space depends on ambient temperature and various characteristics of the enclosure [e.g., platform; (18)]. Therefore, in the Good Housing principle all of these aspects should be taken into account. Some animal-based measures related to comfort around resting are fully stretched, wet, and dirty animals. Most of the farms assessed had at least 20% of the animals fully stretched, and the maximum achieved was 65% of the animals with this posture, so the results for this parameter were quite good. In addition, the values for wet and dirty animals were excellent in most of the farms. In fact, wire-mesh flooring allows easy passage of manure and urine, is easily cleaned and sanitized, and is associated with lower rates of gastrointestinal disease and better air quality in commercial production systems (17). Nevertheless, certain types of wire-mesh flooring may increase the prevalence of pododermatitis in adult rabbits and the routine provision of a slatted plastic resting mat improves animal comfort and reduces the occurrence of this problem (17). In the present study, 90% of the farms were providing these resting mats to all their animals, which is a good sign concerning the quality of the farms assessed, and follows the progress already found by Rosell and de la Fuente (19), where their use in Spanish farms is described to have increased from 28% in 2001 to 75% in 2012. On the other hand, injuries inflicted on the kits by the doe can be minimized by enabling the doe to retreat from kits once they begin leaving the nest box, for instance by providing a platform (20). Platforms enable the possibility for more movement and can improve bone quality by enabling weight-bearing activity [e.g., jumping; (17)]. However, none of the farms provided these platforms to the animals. The farmers justified this because, unlike the provision of resting mats, the inclusion of platforms implies an enormous investment due to dramatic modifications needed in the cages and facilities. This could be an example of how a label system that could provide some extra added value to the final product could be used for investments to improve the welfare of animals, as this one could be an area to prioritize.

The respiratory tract of rabbits is irritated by fine dust in the air (21), so dust levels should not be too high. Nonetheless, there is no literature showing which dust levels are acceptable or not. For other species, Welfare Quality applies a dust sheet test which is a simple procedure indicating the amount of dust in the air (16), and applying the same methodology, in this study, 100% of the farms obtained an excellent rating for this parameter. In fact, the subjective assessment of the auditor agreed with this result, as the farms assessed were in all cases well-ventilated. Another aspect where all of the farms obtained an excellent was the quality of nesting material. Young kits, particularly those <2 weeks of age, have a very limited ability to thermoregulate, so properly bedded nest boxes provide warmth, minimizing chilling and mortality (17). Finally, light quality is another important aspect to consider when comfort around resting is considered. Lighting should provide uniform illumination and permit effective observation of rabbits. A light intensity of 30–50 lux at the rabbit level is necessary to enable mature rabbits to investigate their surroundings, have visual contact with other rabbits, and show active behaviors (11). Continuous lighting (i.e., no dark period in a 24-h cycle) negatively impacts welfare and health, and a natural light-dark pattern enables the rabbit to apply its natural rest-activity rhythm (17). Although light intensity and light-dark pattern was correct in 87% of the farms assessed, four farms failed in this parameter, all four of them for providing fewer than 8 h of enough light to the animals.

The second criterion to consider inside good housing is thermal comfort. When ambient temperatures exceed 25°C, rabbits begins to be at risk of heat stress, which may be indicated by decreased feed intake, increased water intake, open-mouthed panting with the head extended backwards, salivation, and ears fully upright and expanded with prominent blood vessels (21). When ambient temperatures exceed 35°C, rabbits can no longer regulate body temperature and are at significant risk of hyperthermia and heat stroke (22). However, the effective environmental temperature (i.e., the temperature that animals actually feel) may differ by several degrees from that measured in the overall barn and depends on several factors, such as air speed and temperature, relative humidity, flooring and cage/pen type, bedding, single or group housing, and the animal's stage of production and health status (17). As it is not possible to always register all of these variables, in the protocol a range is proposed of acceptable temperatures that can reach 28°C, but the score obtained can be corrected if the animals are seen panting. In addition, data records of maximum and minimum daily temperatures in the long-term are requested. None of the farms assessed had any animal showing panting or shivering, but one-third of the farms failed in the score because of the lack of data, so this is another important point for future improvements.

Ease of movement mainly considers height of the cage and space allowance. Space allowance affects a rabbit's ability to perform behaviors important to the species (e.g., grooming, hopping, jumping), and to adopt normal resting postures (ventral and lateral) and sitting postures [sitting upright or with all four legs on the ground; (17, 18)]. Providing an area within the cage/pen with a minimum height of 40 cm promotes the expression of natural behavior and reduces the risk of ear lesions (17). In the present study, only one farm had a height of 40 cm, and only 30% reached 38 cm. In terms of space allowance, a breeding rabbit toward the end of pregnancy (4 kg−5 kg live weight) would need a cage with a minimum of 3,500 cm^2 according to the EFSA (11), but only 50% of the farms in the present study had at least these 3,500 cm^2, with some farms having 4,000 cm^2. Although Mirabito et al. (23) did not observe any significant difference in reproductive time or budget time of reproducing does kept in cages with different available surfaces (about 3,400, 4,500, and 5,900 cm^2), Dresher (24) showed a great reduction in abnormal skeletal developments in cages of 3,500 cm^2 as compared to 2,400 cm^2. Therefore, a strategy, probably again with a label giving an added value to pay for the investment necessary, is needed to encourage all farms to be able to arrive to at least the 3,500 cm^2 and 40 cm of height.

Good Health

This principle includes absence of injuries, diseases, and pain induced by management. Rabbits should not have any skin damage or wounds. Wounds can be caused by inadequate equipment (e.g., sharp parts of cages), or by mutilative or aggressive behavior of other rabbits. In general, the results obtained in the present study were very positive in this respect, with only a few farms ($n = 4$) with some animals with wounds on the body, and none with problems in the ears or gait score. Pododermatitis should be considered apart. The condition begins with localized hair loss and callous formation on the footpad and the hind feet. It progresses to cracked and open calluses and is most severe when open wounds or ulcers have formed (17). Although 90% of the farms had rest mats, only 48% of the farms obtained an excellent for this parameter, and 30% of the farms had more than 8% of the animals with severe cases, so other strategies should be considered to reduce the incidence of this painful problem. Routine maintenance of facilities and timely replacement of cages/pens before their condition deteriorates helps prevent rabbits from becoming injured, but at least two cages in 27% of the farms assessed were considered dangerous for the animals, so this is an important point to consider, and this justifies that 30% of the total score of the criterion could depend on this fact independently of the values found in other animal-based parameters.

To check if rabbits suffer from disease, they can be checked for a number of clinical signs that are indicative of health problems, like coughing and sneezing, nasal and ocular discharge (25); mange, dermatitis (26), or diarrhea (25). In the present study, coughing, sneezing, ocular discharge, and neck torsion were not found. In addition, just one farm had animals affected by skin abnormalities, such as dermatitis, another one with animals affected by enteropathy, and another one with problems of mange. In fact, the two health problems most seen were nasal discharge, with 17% of the farms with more than 4% of the animals affected (and a maximum of 12%), and diarrhea, the most predominant problem, with 50% of the farms with more than 4% of the animals affected (and a maximum of 20%). However, globally, the presence of diseases in the assessed farms could be considered low. On the other hand, mortality is an important indicator of herd health to monitor on a farm (17, 27, 28). In fact, in breeding rabbits, mortality is often due to infectious causes (17), and a reduction in this indicator represents an improvement in animal health (28). When monthly mortality in breeding does and bucks due to adverse health issues and injury exceeds 5%, it should be considered as an alarm signal (17). In this respect, 50% of the farms assessed had mortalities in the last 3 months below 5%, being considered acceptable values, and 17% of the farms had <3% of mortality, being considered excellent values. Therefore, globally, the values of mortality confirmed the general good health found the day of the visit. Nevertheless, another 17% did not have data on mortality. This should be considered unacceptable and leave the farmer who is not providing these numbers out of any certification system. In any case, mortality records should be interpreted in conjunction with culling records. In fact, for all conditions affecting rabbit health and welfare, early recognition and prompt treatment or euthanasia are essential to minimize animal pain and distress (17). Thus, in the present protocol, the percentage of culling is also assessed, and it is considered that, ideally, the animals culled in a farm should be more than those deaths without human intervention, the total numbers being, of course, as low as possible. According to Rosell and de la Fuente (29), the median monthly removal risk in does in Spain from 2000 to 2005 was 9.3%, with 3.4% dead and 5.7% culled. Therefore, in a period of 3 months, as assessed in the present protocol, numbers would be around 10% of mortality + 15% culling, so most of the farms assessed in the present study present a clear improvement in these numbers (**Figure 2**). However, 40% of the farms had higher mortality rates than culling rates, so emergency killing is still something to be improved in an important number of farms. For does, and related to culling, the percentage of replacement may be a good indicator of health. According to Marai et al. (30), percentage of replacement of does varies between 70 and 160% per year. In the present study, 13% of the farms ranged from 35 to 70%, and all of them showed an excellent result in relation to the presence of diseases. In fact, two of them, Farms 7 and 27, obtained the highest overall scores of the 30 farms, 74 and 73 points, respectively. Nonetheless, to have good values in this parameter is not a guarantee of high final scores, as Farm 3, with a replacement rate of 35%, obtained a final score of 54 points (**Table 6**), so one point below the objective of the 55 points. In this case, a lower score for the Good Housing principle in comparison to other farms would be the cause. In fact, it was one of the only three farms where rest mats were not present for 100% of the animals the day of the visit (with 8% of animals with severe pododermatitis) and, as in other farms, no data were taken for temperatures. Therefore, this is a good example of how animal welfare assessment should comprise a combination of different measures instead of a single indicator. The fact that 70% of the farms had a replacement of more than 110% is related, as well, to the high rhythms of intensification of most of the rabbit farms. Actually, only 13% of the farms leave at least 49 days between parturitions in does, with most of the farms leaving just 42 days. This is another point that could be improved with a label giving added value to the product.

While milk production varies between does, daily milk production typically peaks toward the end of the third week of lactation and then drops rapidly (21), coinciding with the period in which kits' intake of solid feed increases. In natural conditions, if the doe is not pregnant, litter weaning is completed within the fifth and sixth weeks of age (31). This is also an important time to consider welfare in rabbit production, as it has been shown there is an association between later weaning and increased risk of enteritis because of increased stress (17). In fact, long weaning times affect the welfare of the doe due to an increased demand of energy when lactation is combined with gestation (especially in short intervals between parturitions) and the impossibility to escape from offspring in most usual facilities. Therefore, although the objective of the present protocol is to give the highest score to the farms weaning the animals at 35 days of age or later, and this occurred in 63% of the farms, the objective of the protocol should be to balance this parameter with long intervals between parturitions and improved facilities to provide some opportunities to the doe for escaping from the kits, all combined with lower replacement rates and good general health status. Finally, another aspect to consider in the absence of disease criterion is the general cleanliness of the facilities, scored with an excellent in 100% of the farms assessed. Related to cleanliness, rodents and insects are recognized as carriers of many diseases. As the control of rodents should be a legal requirement, the protocol is centered on the presence of flies. In this respect, 73% of the farms had neither flies nor fly eggs present in the farm, while the rest had both flies and their eggs present, being considered a risk for diseases and consequently penalized in this criterion.

Pain induced by management includes two main aspects, the mutilations performed on the animals for identification and how emergency killing or euthanasia is performed. Euthanasia is defined as the "ending of the life of an individual animal in a way that minimizes or eliminates pain and distress" (32). It is characterized by rapid, irreversible unconsciousness followed by prompt death (33). Euthanasia is an important aspect of animal welfare. Allowing a sick or injured rabbit to linger unnecessarily is unacceptable. Any euthanasia method must result in rapid loss of consciousness followed by death without the animal regaining consciousness (33). Neck dislocation was the system used by 93% of the farmers for does and bucks in the present study. According to the recommendations of EFSA (34), cervical dislocation is considered a killing method and therefore it should only be applied on unconscious animals. In addition, the hazards related to cervical dislocation include "manual restraint" (leading to pain and fear) and "incorrect application" [leading to the absence of unconsciousness, pain, fear, and distress; (34)]. For these reasons, in the present protocol, the system is not considered as correct. Certainly, this is one critical point that needs to be improved in the rabbit farms assessed. For identification, ear-marks (metal or plastic), microchips, or tattoos can be used. There are countries where ear-marks are not allowed, for being potentially painful to the rabbit (17) or causing injury by being caught on the cage wire. Consequently, in the protocol it is asked to not use this system, which agrees with the practice of 97% of the farms assessed, which use tattoos for identification. However, Keating et al. (35) described an acutely painful procedure for rabbits related to tattooing the ears and suggest the use of anesthesia to mitigate the associated pain. Therefore, to ask for the use of anesthesia, if tattooing is needed for identification, could be an appropriate refinement of the protocol.

Appropriate Behavior

Managing territoriality and associated aggression in pair- or group-housed does is difficult. Rates of doe injuries and kit mortality are typically higher in pair and group systems (17). This was not a problem in the farms assessed in the present study, as all adult animals were housed in single cages. However, when cages are too close to each other, dominant animals can try to bite the adjacent subordinates when lying by the cage wall. This would have a higher risk of injuries, prevents a correct resting behavior in the subordinate animal, and induces a higher alert level in the dominant one, so it is suggested to have a good separation between cages or provide solid walls to prevent contact between individually housed rabbits (36). In the present study, although no animals were found biting other animals, cages did not have solid walls and the distance between cages was insufficient, so 0 points were obtained in all cases for the social behavior criterion. However, a recent paper (37) describes that adult rabbits are better in pairs than alone. Although this study is done in neutered rabbits in outdoor conditions with low temperatures where huddling was needed to maintain body temperature and in just 45 individuals from a rabbit-only rescue center, this should be considered in further studies.

Both in growing and adult animals, stereotypes, which is abnormal behavior repeated obsessively without apparent aim, have often been described (38, 39). Stereotypes and abnormal behavior are indicators of reduced welfare in rabbits. These behaviors can be head shaking, swaying, wire gnawing, wall pawing, and over-grooming (11, 40–42). In the present study, none of the animals were observed showing these behaviors. Although the methodology, 2 min of observation per animal, was tested previously in other farms with positive results (presence of stereotypes) and the time dedicated to assess abnormal behaviors is double that used in other species, such as pigs, for the same purpose (16), it cannot be discarded that some adjustment could be necessary to the methodology to increase its sensitivity. In any case, rabbits perform fewer abnormal behaviors (e.g., oral stereotypies, cage biting, or manipulation) when provided with enrichment material (17), so this is another important point to consider in the protocol. Examples of enrichment are hard wood-gnawing of blocks or sticks, hay, straw or litter (for chewing or manipulation), grass or hay in any form, tubes/tunnels, and mirrors (17). Nevertheless, none of the farms assessed was using any kind of enrichment for their rabbits, so this is clearly another gap that the rabbit producers need to address. An important exception to the absence of any type of enrichment material is the provision of nesting material for does 24 h prior to kindling. If insufficient nesting material is present, the doe cannot perform her natural nest building behavior (43). A variety of bedding materials may be used, including rabbit hair, hay, straw, shredded paper, and wood shavings, but in any case, it should be dry and dust free. In this case, nesting material for the doe was found to be proper (in quantity and quality) in all farms assessed.

In natural conditions, after kindling and attending to the newborn kits, the doe leaves the nest, closes it up, and comes back only to suckle the kits. According to Trocino and Xiccato (31), suckling takes place once a day, usually after sunset, and lasts a few minutes (two to five), during which the kits ingest a high quantity of nutritive substance and energy enough for rapid development and growth. However, Hoy et al. (44) described that only 56% of the does free to enter cages at will, in fact, nursed their kits only once a day, whereas 40% nursed twice or more often during the day and 4% did not nurse at all. Most of the rabbit farmers, and 100% of the ones assessed in this study, reproduce this behavior in their farms by means of a controlled lactation. This is to open the nest for allowing the doe to visit her kits just once a day and for a few minutes. The advantages of controlled lactation are a reduction in kit mortality due to crushing and higher kit weight homogeneity (45). In addition, it helps to confirm that nursing is occurring. Further, not keeping to a regular timetable and leaving the doe waiting to access the nest negatively compromises her welfare through increased anxiety and likely physical discomfort caused by delayed nursing opportunity. For this reason, the protocol considers at which hour this operation is to be carried out every day or whether the nest is freely accessible to the doe throughout the 24 h. In addition, it might be appropriate to consider whether providing access to the nest twice a day should score higher than once a day. The few differences among farms, especially because of the absence of abnormal behaviors and the absence of enrichment material, produced a high homogeneity in the final score of the whole criterion, with 60 points in all cases.

A good human-animal relationship promotes rabbit welfare. With proper handling, rabbits experience less stress and fear, and the risk of injury to the animals and handlers is greatly reduced (11). In other species, the human-animal relationship is assessed by means of an approach test (16). The approach test of the present protocol is based on those proposed by Hansen and Moller (46) for minks maintained in cages but also considering the lack of aggressivity of rabbits against the stick. The range of animals touching the stick was from 0% in the worst farm to 20% in the best one. Although 30 farms are not probably enough to have a complete picture of the situation, it is possible that in the future the thresholds of what is acceptable and excellent could be adjusted to lower percentages to increase the capacity of discrimination of the parameter. Another important point to consider in this criterion is training. Management practices have a significant impact on animal health, welfare, and productivity (47). In addition, training and knowledge development about rabbit welfare and care should be an ongoing process. Nevertheless, 57% of the farms assessed did not have any person trained in animal welfare, so this is again a critical point that producer associations should try to solve. Among the actions that may be adopted to improve human-animal relationships, early manipulation has been shown to provide positive results, especially if it is applied during a sensitive period in the first week post-partum and near the time of nursing, due to a general increase in arousal that occurs at this time (48). For this reason, it is asked in the protocol if the kits are touched at least once a day during the first week of age. All of the farms assessed were performing a controlled lactation and ensuring that all kits were taking in milk (see Good Feeding, above) by holding the animals gently and checking the abdomen, so an excellent was obtained for this parameter. Globally, this was the criterion with the most variability within the three of the Appropriate Behavior principle.

Global Assessment

As commented previously, the farms assessed in the present study were asked to obtain a minimum of 55 points in the global score and they had the opportunity to check the assessment protocol prior to being audited under a voluntary basis. Therefore, few farms, even none, should be expected to score below 55 points. However, this occurred in five farms. Three of them were actually very close to 55 points (52 and 54 points) and the other two not (41 and 48 points). In addition, 6 other farms were between 55 and 60 points (**Table 6**). In all of these cases, it is encouraged to carefully review the critical points found during the assessment to improve the final score. In the four farms below 55 points this is mandatory, but for the other farms it can be important as well for not having problems in the future. In fact, as mentioned previously, the disease results were fine, in general, in the assessed farms, so these farms so close to 55 points are at risk if, in the next assessment, the health score, due to punctual problems, has a worse result. In addition, the rest of the farms are still far from the 100 points or even the excellent, due to not dealing with health issues punctually in order to reduce suffering. So in all cases there are opportunities for improvement. Therefore, the protocol can be used as well as a tool for identifying gaps and planning future investments. Good Housing and Appropriate Behavior are the two principles with the lowest weights in the final score (15 and 20%, respectively), and at the same time they are the principles with the lowest variability. In the first case it is because of a very high score, and in the second because of a very low score. Although a future version of the protocol could try to add more variability to the first principle and penalize in the final value more if the last one has such low scores, the results just show how, for years, rabbit producers have been very focused on feeding needs and very few on behavioral needs.

CONCLUSIONS

In general, most of the farms obtained a good overall score, the maximum found being 73 points. Nevertheless, none of the farms obtained an excellent, and four farms were scored below the 55 points required. The Good Feeding principle obtained the highest score, reaching an excellent in all farms, and Appropriate Behavior the lowest one, with values ranging from 21 to 41 points out of 100. In general, the main problems found were absence of platforms, low space allowance and height of the cage, inappropriate system for performing emergency killing, insufficient protection of does from other adjacent does when housed individually, absence of enrichment material, and, in some cases, the lack of temperature data records, high replacement rates, and even lack of mortality rate data records.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, but maintaining blinded the farms of origin.

ETHICS STATEMENT

Ethical review and approval was not required for the animal study because the study is focused in the assessment of farm animal welfare in commercial farms, so no interventions of any kind were carried out on the animals.

AUTHOR CONTRIBUTIONS

AD was responsible for the development of the protocol, collecting the data, writing the paper, and training the assessor in using the protocol. JP and XM were collaborators in developing the protocol and testing it in commercial farms during its development to refine the parameters. All authors contributed to the article and approved the submitted version.

FUNDING

Most of this project was funded with funds from the research institute (IRTA) and a small part with the participation of private companies not related to rabbit production.

ACKNOWLEDGMENTS

Thanks are given to Pau Batchelli for his work collecting the data used in this paper and all the experts that collaborated in the development of the protocol, with special mention to Jordi Garcia and Felix Freixer for providing as well their farms for testing the firsts versions of the protocol. The 30 farms visited during the study are acknowledged too.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2020.00445/full#supplementary-material

REFERENCES

1. FAOSTAT. *Food and Agriculture Data*. Food and Agriculture Organization of the United Nations (2020). Available online at: http://www.fao.org/faostat/en/#data
2. Broom DM. The scientific assessment of animal welfare. *Appl Anim Behav Sci*. (1988) 20:5–19. doi: 10.1016/0168-1591(88)90122-0
3. Fraser D, Ritchie JS, Fraser AF. The term 'stress' in a veterinary context. *Brit Veter J*. (1975) 131:653–62. doi: 10.1016/S0007-1935(17)35136-9
4. Simplicio JB, Cervera C, Blas E. Effect of two different diets and temperatures on the growth of meat rabbit. In: *Proceedins of Fourth Congress World Rabbit Science Association* (Budapest). (1988). p. 74–7.
5. Paci G, Bagliacca M, Marzoni M, Liponi GB. Consumo giornaliero e CUDa di conigli sottoposti a differenti temperature. *Ann Facol Med Vet Pisa*. (1993) XLVI:377–85.
6. Good RA, Lorenz E. Nutrition and cellular immunity. *Int J Immunopharma*. (1992) 14:361–6. doi: 10.1016/0192-0561(92)90165-H
7. Kasa IW, Thwaites CJ. Semen quality in bucks exposed to 34°C for 8 hours on either 1 or 5 days. *J Appl Rabbit Res*. (1992) 15:560–8.
8. Marai IFM, Rashwan AA. Rabbits behavioural response to climatic and managerial conditions–a review. *Archiv Tierzucht Dummerstorf*. (2004) 47:469–82. doi: 10.5194/aab-47-469-2004
9. Lebas F, Coudert P, Rouvier R, de Rochambeau H. *The Rabbit. Husbandry, Health and Production*. Animal Production and Health Series. Rome: FAO (1986).
10. Dalmau A, Catanese B, Rafel O, Rodriguez P, Fuentes C, Llonch P, et al. Effect of high temperatures on breeding rabbit behaviour. *Anim Prod Sci*. (2015) 55:1207–14. doi: 10.1071/AN13440
11. EFSA. The impact of the current housing and husbandry systems on the health and welfare of farmed domestic rabbits. *EFSA J*. (2005) 267, 1–31. Available online at:http://www.asfc-lapin.com/Docs/Activite/T-ronde-2011/Fichiers-pdf/Rapport-EFSA.pdf
12. Blokhuis HJ, Veissier I, Miere M, Jones RB. The Welfare Quality project and beyond: safeguarding animal well-being. *Acta Agric Scand Sect A Anim Sci*. (2010) 60:129–40. doi: 10.1080/09064702.2010.523480
13. De Jong IC, Reuvekamp BFJ, Rommers JM. *A Welfare Assessment Protocol for Commercially Housed Rabbits*. Report 532. Wageningen: Wageningen UR Livestock Research (2011).
14. Bonanno A, Mazza F, Di Grigoli A, Alicata ML. Body condition score and related productive responses in does. In: *Proceedings of 9th World Rabbit Congress*, Verona (2008). p. 297–301.
15. Popescu S, Diujan EA, Borda C, Mahdy CE. Welfare assessment of farmed rabbits housed in indoor and outdoor cages. *Anim Sci Biotechnol*. (2013) 46:200–5.
16. Welfare Quality®. *Welfare Assessment Protocol for Poultry*. Lelystad: Welfare Quality® Consortium (2009).
17. National Farm Animal Care Council. *Code of Practice for the Care and Handling of Rabbits: Review of Scientific Research on Priority Issues*. Lacombe, AB: National Farm Animal Care Council (2018). Available online at: https://www.nfacc.ca/pdfs/codes/rabbit_code_of_practice.pdf
18. Giersberg MF, Kemper N, Fels M. Planimetric measures of floor space covered by fattening rabbits and breeding does in different body postures and weight classes. *Livest Sci*. (2015) 177:142–50. doi: 10.1016/j.livsci.2015.04.010
19. Rosell JM, de la Fuente LF. Assessing ulcerative pododermatitis of breeding rabbits. *Animals*. (2013) 3:318–25. doi: 10.3390/ani3020318
20. Verga M, Luzi F, Szendro Z. Behavior of growing rabbits. In: Maertens L, Coudert P, editors. *Recent Advances in Rabbit Sciences*. Melle: ILVO (2006). p. 91–7.

21. Lebas F, Coudert P, de Rochambeau H, Thébault RG. *The Rabbit–Husbandry, Health and Production*. Food and Agriculture Organization of the United Nations (1997). Available online at: http://www.fao.org/docrep/t1690e/t1690e00.htm#Contents
22. Marai IFM, Habeeb AAM, Gad AE. Rabbits' productive, reproductive and physiological traits as affected by heat stress: a review. *Livest Prod Sci.* (2002) 78:71–90. doi: 10.1016/S0301-6226(02)00091-X
23. Mirabito L, Dumont F, Galliot P, Souchet C. Logement collectif des lapines reproductrices: Conséquences sur le comportement. In: *Proceedings: 11èmes Journées Recherche Cunicole*, Paris (2005). p. 57–60.
24. Dresher B. Deformations of vertebral column in breeding rabbits. In: *Proceedings: 6th World Rabbit Congress*, Toulouse (1996). p. 417–21.
25. Peeters JE. Ademhalings- en spijsverteringsstroornissen in de industriele slachtkonijnenhouderij. *Diergeneesk Memorand.* (1989) 3:89–136.
26. Meek B. Huidproblemen bij het konijn. Deel 1: Besmettelijke huidproblemen. *Kleind Mag.* (2011) 125:38–40.
27. Rosell JM, de la Fuente LF. On-farm causes of mortality in female rabbits. In: *Proceedings of the 10th World Rabbit Congress* (Sharm El-sheikh). (2012). p. 1147–50.
28. Depoorter P, Van Huffel X, Diricks H, Imberechts H, Dewulf J, Berkvens D, et al. Measuring general animal health status. *Prev Vet Med.* (2015) 118:341–50. doi: 10.1016/j.prevetmed.2014.12.008
29. Rosell JM, de la Fuente LF. Culling and mortality in breeding rabbits. *Prev Vet Med.* (2009) 88:120–7. doi: 10.1016/j.prevetmed.2008.08.003
30. Marai IFM, Askar AA, McKroskey RA, Tena E. Replacement in rabbit herds: a review. *Trop Subtrop Agroecosyst.* (2010) 12:431–44.
31. Trocino A, Xiccato G. Animal welfare in reared rabbits: a review with emphasis on housing systems. *World Rabbit Sci.* (2006) 14:77–93. doi: 10.4995/wrs.2006.553
32. American Veterinary Medical Association. *AVMA Guidelines for the Euthanasia of Animals: 2013 Edition*. Schaumburg, IL: American Veterinary Medical Association (2013).
33. Canadian Veterinary Medical Association. *Euthanasia–Position Statement*. (2014). Available online at: www.canadianveterinarians.net/documents/euthanasia
34. EFSA. Scientific opinion concerning the killing of rabbits for purposes other than slaughter. *EFSA J.* (2020) 18:5943. doi: 10.2903/j.efsa.2020.5943
35. Keating SCJ, Thomas AA, Flecknell PA, Leach MC. Evaluation of EMLA cream for preventing pain during tattooing of rabbits: changes in physiological, behavioural, and facial expression. *PLoS ONE.* (2012) 7:e44437. doi: 10.1371/journal.pone.0044437
36. RDA. Rapport van aanbevelingen welzijnsproblematiek in de bedrijfsmatige konijnenhouderij. In: Baars GMJ, Bokkers EAM, van der Hage MH, van Haren HIJ, Hurkmans FH, de Jong M, et al., editors. *Leden van de Werkgroep bedrijfsmatig houden van konijnen*. Lelystad: RDA (1997). p. 45.
37. Burn CC, Shields P. Do rabbits need each other? Effects of single versus paired housing on rabbit body temperature and behaviour in a UK shelter. *Anim Welfare.* (2020) 29:209–19. doi: 10.7120/09627286.29.2.209
38. Verga M, Carenzi C. *Il Comportamento degli Animali Domestici*. Bologna: Edagricole (1981).
39. Lawrence AB, Rushen J. *Stereotypic Animal Behaviour: Fundamentals and Applications to Welfare*. Wallingford: CAB International (1993).
40. Chu L, Garner JP, Mench JA. A behavioral comparison of New Zealand White rabbits (*Oryctolagus cuniculus*) housed individually or in pairs in conventional laboratory cages. *Appl Anim Behav Sci.* (2004) 85:121–39. doi: 10.1016/j.applanim.2003.09.011
41. Gunn D, Morton DB. Inventory of the behavior of New Zealand White rabbits in laboratory cages. *Appl Anim Behav Sci.* (1995) 45:277–92. doi: 10.1016/0168-1591(95)00627-5
42. Lidfors L, Edstrom T, Lindberg L. The welfare of laboratory rabbits. *Anim Welfare.* (2007) 2:211–43. doi: 10.1007/978-1-4020-2271-5_10
43. Stauffacher M, Baumans V. Assessment of animal housing standards for rabbit in a research setting. In: *International Workshop on the Development of Science-Based Guidelines for Laboratory Animal Care*. (2003). p. 86–90.
44. Hoy ST, Seitz K, Selzer D, Schüddemage M. Nursing behaviour of domesticated and wild rabbit does under different keeping conditions. In: *Proceedings of 7th World Rabbit Congress*. Valencia (2000). p. 537–43.
45. Coureaud G, Shaal B, Rochon JJ, Richard F, Bohec V. Le contrôle de l'accès au nid chez le lapine: conséquences sur la mortalité des lapereaux. In: *Proceedings: 7èmes Journées Recherche Cunicole*. Lyon (1998). p. 245–8.
46. Hansen SW, Moller SH. The application of a temperament test on-farm selection of mink. *Acta Agric Scand Sect A Anim Sci.* (2001) 51:93–8. doi: 10.1080/090647001316923144
47. Rushen J, Passillé AM. The importance of good stockmanship and its benefits for the animals. In: Grandin T, editor. *Improving Animal Welfare: A Practical Approach*. Cambridge, MA: CABI International (2010). p. 50–63.
48. Bilkó A, Altbäcker V. Regular handling early in the nursing period eliminates fear responses toward human beings in wild and domestic rabbits. *Dev Psychobiol.* (1999) 36:78–87. doi: 10.1002/(SICI)1098-2302(200001)36:1<78::AID-DEV8>3.0.CO;2-5

Conflict of Interest: The data was obtained from farms interested in being certified on animal welfare within a certification scheme owned by IRTA. Therefore, although it exists as a commercial relationship, the farmers were just subjected to the audit with no possibilities of any other intervention in the study. The authors therefore 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.

frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 03 September 2020
doi: 10.3389/fvets.2020.00607

Variations in the Behavior of Pigs During an Open Field and Novel Object Test

Amy Haigh[1*†], *Jen-Yun Chou*[1,2,3] *and Keelin O'Driscoll*[1]

[1] Pig Development Department, Animal and Grassland Research and Innovation Centre, Teagasc, Moorepark, Fermoy, Ireland, [2] Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, United Kingdom, [3] Animal and Veterinary Sciences Research Group, Roslin Institute Building, Scotland's Rural College (SRUC), Easter Bush, United Kingdom

OPEN ACCESS

Edited by:
Edward Narayan,
The University of Queensland, Australia

Reviewed by:
Anna Valros,
University of Helsinki, Finland
Manja Zupan Šemrov,
University of Ljubljana, Slovenia

***Correspondence:**
Amy Haigh
amyjohaigh@yahoo.com

†Present address:
Amy Haigh,
School of Biology and Environmental Sciences, University College Dublin, Dublin, Ireland

Specialty section:
This article was submitted to Animal Behavior and Welfare, a section of the journal Frontiers in Veterinary Science

Received: 17 April 2020
Accepted: 28 July 2020
Published: 03 September 2020

Citation:
Haigh A, Chou J-Y and O'Driscoll K (2020) Variations in the Behavior of Pigs During an Open Field and Novel Object Test. Front. Vet. Sci. 7:607. doi: 10.3389/fvets.2020.00607

Tail biting is a serious welfare concern in pig production. It not only causes distress for victims, but may occur where pigs are unable to cope, and become biters. An animal's ability to cope with stressful situations may vary between individuals, but the behavioral response could be consistent across different fear eliciting situations. We exposed 75 pigs to open field (OF) and novel object (NO) tests at 14 weeks of age. Within each pen of pigs ($n = 16$ pens, 55 pigs/pen), 6 pigs were selected for testing using the following criteria: 3 pigs that had severe bite wounds (BITTEN), 1 confirmed biter (BITER), 1 pig which could be easily approached and trained to provide a saliva sample (BOLD) and 1 pig which was extremely evasive, and was unable to be trained to willingly provide a saliva sample (SHY). Given that responses may be consistent in different scenarios, we hypothesized that SHY pigs would display more characteristics of a fear response (i.e., less movement in the open field, more time spent by the door, and longer latency to approach the novel object) than human BOLD pigs. We also hypothesized that BITTEN pigs would behave similarly to SHY and BITERS similarly to BOLD. The BOLD and BITER pigs spent more time exploring ($P < 0.05$) and less time by the door ($P < 0.01$) than the BITTEN and SHY pigs. Although there was an overall increase in cortisol level from before to after the tests ($P < 0.001$), this was only significant for BITTEN ($P < 0.001$) and SHY ($P < 0.05$) pigs. Therefore, as hypothesized, for several measures, BOLD, and BITER pigs behaved similarly, and differently to SHY and BITTEN. However, the low sample size potentially meant that for several measures, although numeric differences were in the direction hypothesized, there were no statistical differences. Further work in which a greater number of BITER pigs were included in the sample, may elucidate our hypotheses more clearly, as to whether responses to fear tests in pigs could be associated with the likelihood of being a tail biter, or victim.

Keywords: tail biting, fear, reactive, proactive, cortisol

INTRODUCTION

Tail biting is one of the most significant problems in pig production, as it causes pain and distress for victims (1) and financial loss for producers (2). The problem has a multifactorial etiology, with damaging biting behavior thought to occur through three main mechanisms: "two-stage," which consists of mild chewing escalating to severe biting (thought to arise from a lack of manipulable material), "sudden-forceful," which is associated with competition for access to resources, and "obsessive" whereby the causal factors are not known, but the biter moves from pig to pig, compulsively biting tails (3).

Open Field (OF) and Novel Object (NO) tests are often used to study fearfulness in pigs, and indeed have previously been used to compare the responses of tail biters, tail bite victims, and non-affected animals (4–6). In general, these tests have shown that biters and/or victims react differently to "control" pigs both behaviourally (biter pigs less likely to approach a NO (4) more time lying and walking, less standing still, lower latency to touch the object (6), and physiologically (e.g., dysfunctional autonomic regulation (6), in ways which indicate that the animal is fearful. However, it is unclear as to whether these differences relate to a generalized fearfulness, or whether they are specific to exposure to this type of test. Moreover, in these studies, the control pigs were pigs which were not affected by biting, rather than having been selected because they have a propensity to react in a fearful or confident manner to novel or fear eliciting situations.

Pigs can respond differently depending on the type of behavioral test which is applied; this is because tests of fearfulness can comprise of varying levels of fear eliciting stimuli, such as novelty, physical proximity, movement, intensity, suddenness, and duration (7). Pig level characteristics such as their age and sex can also influence the response (8), as well as their personality, which is generally considered to consist of individual differences which are consistent over time and situations (9). Thus, pigs which respond fearfully or with confidence during any particular situation, may not display the same characteristics in another.

This experiment was carried out as a component of a larger study comparing compressed straw blocks or plastic toys as enrichment for pigs from weaning to finish (10). The aim of the work was to initially investigate whether pigs which were human BOLD in their home pen (could be approached/would willingly approach people, and could easily be trained for cortisol collection) would respond in a less fearful manner to both an open field (OF) and novel object (NO) test, than pigs which were human SHY in the home pen (couldn't be approached/wouldn't willingly approach people and couldn't easily be trained for cortisol collection but would consistently run away) (i.e., was their behavior consistent across the two scenarios). We then investigated whether pigs which were either tail BITERs or BITTEN would respond differently to the tests. Given that tail biting is associated with increased stress levels in victims (11) we hypothesized that BITTEN pigs would display behavior similar to SHY pigs (i.e., a reduced ability to respond to a fear eliciting situation), whereas BITER pigs would respond similarly to BOLD, as the novel situation would provide an outlet for exploratory behavior (12).

METHODS

The study was conducted on a commercial farm, ∼10 miles from the Animal and Grassland Research and Innovation Center, Teagasc Moorepark Fermoy, Co. Cork, Ireland. Prior to commencing the experiment approval was obtained by Teagasc Animal Ethics Committee (TAEC89/2015).

Animals and Treatments

A total of 880 pigs (Terminal line PIC, born from Large White X Landrace sows) which were born in two replicates 7 weeks apart (440 per replicate) were used in the experiment. Approximately 75% of the tail was docked at 3 days of age (as per veterinary recommendation at the farm), and males were kept intact. Piglets were weighed and individually ear tagged just prior to weaning and split into 16 groups of 55 piglets on the basis of sex and weight. Piglets of the same sex and from the same litters were kept together, to minimize stress due to re-mixing at weaning. Only piglets in excess of 5 kg and devoid of any injury were included in the experiment. Piglets were weaned at 28 days of age, assigned to treatment, and were managed in the same group until the end of the 2nd stage weaner accommodation.

Details of the experiment are published in the paper comparing compressed straw blocks and plastic toys with regard to damaging behavior, by Haigh et al. (10). In brief, the experiment followed a 2 × 2 factorial design with sex (male or female; $n = 8$ pens each) and enrichment (compressed straw (STRAW) or plastic hanging toys (TOY); $n = 8$ pens each) as the main factors. Both types of enrichment (STRAW or TOY) were provided at an allowance of two devices per pen (i.e., ∼27 pigs per device). The straw blocks were cylindrical (∼3 inches diameter), and provided in a dispenser (a plastic cylinder (Medinova, Italy) with an open bottom into which the straw block fitted snugly) attached to the wall of the pen. The straw block was supported by a metal bar suspended under the cylinder, so that ∼4 inches of straw were exposed to the pigs. These cylinders were 30.6 inches in length, and were placed 8 inches from the floor in the 1st stage and 19 inches in the 2nd stage weaner pens, and straw was provided continuously.

In the TOY treatment, pigs were provided with a different type of toy in the 1st stage and 2nd stage weaner pens, appropriate to the age of the pig. In the 1st stage the toy consisted of a Porchichew (Ketchum, U.K) suspended at two points in the center area of the pen. In the 2nd stage weaner pens pigs were provided with a rubber hanging "Easyfix" chew toy (Easyfix rubber products, Canada), and a hanging plastic barrel, again both suspended in the center area of the pens.

Selection of Pigs for Behavior Tests

Six pigs per pen were selected to take part in the open field (OF) and novel object (NO) tests; 1 pig which were consistently "human bold" (BOLD), 1 pig which were consistently "human-shy" (SHY), 1 pigs which was confirmed as an active tail biter (BITER), and 3 pigs which had severe lesions on either the ear or tail (BITTEN). Pigs were only used that corresponded to one of the categories i.e., no bitten/biter pigs had also been trained for cortisol collection.

Bold and Shy

Pigs were identified as being BOLD or SHY using data collected as part of the larger study (10). For that study saliva was collected from a subsample of 11 focal pigs from each pen (176 pigs in total) every fortnight from 2 weeks post weaning, to investigate salivary cortisol level. These pigs were selected on the basis of weaning weight; pigs were sorted by weight and every 5th pig selected, so that the full range of weights in the pen were represented. During the first week post-weaning these 11 pigs per pen were habituated to saliva collection by gradually introducing the cotton buds (Salivette, Sarstedt, Wexford, Ireland) used for collection and the experimenter; pigs were to chew on the cotton bud until it was thoroughly moistened (about 30–60 s/sample). As well as the habituation period, four saliva collection days occurred prior to OF and NO testing. If a pig displayed fearful behavior (i.e., ran away, and/or refused to chew on the cotton bud) collection did not occur that week. Pigs that displayed this behavior at least three times prior to behavioral testing were categorized as SHY, and pigs that habituated extremely quickly and gave a good sample voluntarily each time were categorized as BOLD. Within each pen one SHY and one BOLD pig was selected for testing.

Biter Pigs

During the larger study (10), direct behavior observations were conducted each week. The selection of biter pigs for the OF and NO test were therefore based on a total of eight of these observations, 5 during the first weaner stage (9 ± 2, 14 ± 0, 20 ± 2, 28 ± 1, and 38 ± 4 days post weaning) and 3 during the second weaner stage (46 ± 4, 51 ± 5, and 59 ± 4 days post weaning. Observations were carried out four times on each day between 11:30 (hh:mm) and 15:30. Three trained observers collected all the behavior data, balanced across enrichment type and pens. Observation times were also balanced across enrichment type and pens so that observations for each pen were distributed approximately equally across the recording periods. The behavior of the entire group was observed for 5 min. All incidences of harmful and aggressive behavior and play were counted using continuous observation, with tail and ear biting being defined as oral manipulation of the tail or ear of another pig (3). During these observations, the identity of individuals engaged in ear and tail biting was recorded. Over a total of 85 h of observation (∼5 h per pen), a total of 500 records of individuals biting were collected. Within each pen, a BITER was selected that was observed biting other pigs' tails or ears on at least three occasions (Mean-3.8 ± 1.64 (SD), range = 3–9), and more than any other pig in the pen, during these observation sessions. Priority was given to pigs observed biting both ears and tails and observed biting in both the first and second stage indicating it being a consistent trait and not an occasional occurrence, Sniffing around the tail or ear area was not considered biting, the pig had to have the tail or ear in the mouth. Biting was further authenticated by a jolt or flinch in the victim of the biting.

Bitten Pigs

The tails and ears of all pigs were examined individually by the experimenter walking through the pen, on a fortnightly basis, including the week immediately before behavioral testing. This would have been conducted five times prior to the OF and NO, three in the first weaner stage weaner (18 ± 0, 31 ± 4, and 43 ± 4 days post weaning) and two in the second weaner stage (55 ± 6 and 73 ± 9 days post weaning). Tail lesions were scored using the tail lesion scoring system developed by Hunter et al. (13) (0 = no damage, 1 = mild, 2 = moderate, and 3 = severe lesions). Ear lesions were scored using a modified version of the system developed by Haigh et al. (10), and range from 0 (perfect) to 4 (part of the ear missing). Pigs that had a score of 3 or more to either the tail or at least one ear were categorized as bitten pigs. As per standard farm management, all pigs considered unwell or lame were removed to hospital pens and therefore removed from the experiment.

Open Field (OF) and Novel Object (NO) Tests

Testing took place over five consecutive days per replicate when pigs were in 2nd stage weaner pens, ∼70 days after weaning. Test pigs were removed from their home pen, and a saliva sample collected immediately in the corridor outside the pen. The pig was then moved to a waiting area, along with at least one companion pig, immediately adjacent to the test arena. As soon as the preceding pig was finished its test and removed, the pig entered the arena alone and testing began. A pig was never held in the waiting area for more than the entire duration of testing of the preceding pig. Thus, individual pigs were kept outside their home pen for a maximum of 10 min, including collecting saliva collection, immediately prior to the test.

The test arena was an unused room on the commercial farm, which was a maximum of 10 m from the pigs' home pens, with a solid concrete floor and concrete walls, measuring ∼3 × 3 m. The door into the room consisted of a gate 130 cm in height, which was covered by a wooden panel prior to testing. This was so that when the gate was closed the pigs could not see out of the room. The observer stood outside the room and slightly to the left of the gate, which allowed her to see over the top of the gate and have a full view of the room, without entering it. One corner of the room had a series of metal posts installed in the floor diagonally across the corner, so these were also covered with a wooden panel (thus the room was not entirely square). When testing was complete the observer entered the room and collected a second saliva sample, prior to the pig being brought back to its home pen.

The floor of the test arena was divided into nine areas so that the amount of movement through the room during the tests could be assessed. Behavior (**Table 1**) was recorded by a single observer using the Psion Workabout with observer software (Noldus Information Technology, Wageningen, The Netherlands). The OF test began as soon as the pig entered the test arena. After the pig had spent 5 min in the arena, a novel object was presented to the pig in the form of a red yard brush with a rope attached that was lowered over the gate. The pig was then observed for a further 3 min in the presence of this novel

TABLE 1 | Ethogram of behaviors recorded by continuous observation during the open field and novel object tests.

Behavior type	Behavior	Description
State	Stand	Stationary with all four feet on the floor
	Lie	Stationary with body in contact with the floor
	Walk	Moving in a forward or backward direction or turning around at the same location, with head up
	Exploration	Investigating the floor, wall, or object with the rooting disc. See detailed descriptions below
	Explore floor	Investigating the floor by sniffing, nosing, licking, rubbing, or rooting it with the rooting disc. Pig is oriented toward the center of the room. Rooting disc is either in contact or very close to floor surface.
	Explore wall	Investigating the walls of the arena by sniffing, nosing, licking, rubbing, or rooting it with the rooting disc. Rooting disc is either in contact or very close to wall surface
	Self-groom	Scratching or stimulating a part of the body using the fixtures or fittings
	Play	Individual play behavior, including scampering, jumping/running around
	Attention object*	Attention is directed toward object but the pig has not yet touched it
	Withdraw*	Drawing back from object with or without touching it, while attention is still directed toward it
	Explore object*	Investigating the object by sniffing, nosing, licking, rubbing, carrying, throwing, or rooting it with the rooting disc. Rooting disc is either in contact or very close to the object
Event	Low-pitched vocalization	Short or long grunts
	High-pitched vocalization	Screams, squeals, or grunt-squeals
	Elimination	Defecating or urinating
	Jump	Jumping in air or against a wall of the arena

**indicates behaviors which were only recorded during the novel object test.*

object, using the same ethogram as before but with additional behaviors directed toward the object being also recorded. The duration of the behavioral tests was therefore a maximum of 8 min.

The ethogram of behaviors recorded for the OF and NO tests is shown in **Table 1**. For the OF test, behaviors considered for analysis were the amount of time spent walking, performing exploratory behavior (wall and floor), standing still, and the number of vocalizations (low pitched and high pitched). Additionally transitions between the different areas of the room were counted to provide an estimate of the amount of movement around the pen. The percentage time spent in the squares immediately next to the door was also investigated. The number of pigs which eliminated and attempted to escape were also counted.

For the NO test the main focus of interest was on interaction with the novel object. Thus, the behaviors considered for analysis were the time spent directing attention toward the object, interacting with the object, withdrawing from the object, and the latency to interact with it.

Cortisol Analysis

Immediately after cortisol collection the cotton buds were placed in plastic tubes and refrigerated, then centrifuged for 15 min at 3,000 g, and stored at −20°C until analyzed by an enzyme linked immunosorbent assay (Salivary Cortisol Kit, Salimetrics Europe Ltd, Suffolk, U.K). The minimum detectable concentration of cortisol that could be distinguished from 0 was <0.003 mg/dl. The intra- and inter-assay CV's based on controls were 7.2 and 12.9%, respectively.

Data Management

Due to varying levels of tail biting in the pens, in the second replicate it was not possible to select the planned number of pigs within each type from each pen, as the categories were not all mutually exclusive (e.g., BOLD pigs were also BITTEN etc.). These pigs were removed from the analysis. The final number of pigs which were available for analysis within each type are shown in **Table 2** (Total, $n = 75$ pigs).

Statistical Analysis

Data were analyzed using the Statistical Analyses System (SAS, V9.1.3, SAS Institute Inc., 1989). Data were investigated for skew, kurtosis, and outliers before analysis by examination of box and normal distribution plots.

Behavior data were initially explored using correlations (Pearsons and Spearmans Rank, where appropriate) and principle component analysis (PCA; Proc Factor), with the aim of determining whether some of the recorded behaviors clustered and share the same underlying motivation. For the PCA, the numbers of animals which eliminated and jumped were not included due to the an extremely high number of animals eliminating ($n = 69$), and low number jumping ($n = 4$). A varimax rotation was used, as components were considered to be orthogonal. Components were considered meaningful by considering whether the eigenvalue was >1, evaluation of the Scree plot, the proportion of variance explained by components individually and combined, and by considering the interpretation of the output (14).

Data for the OF and NO tests were subsequently analyzed separately using linear mixed models, where residuals confirmed

TABLE 2 | Numbers of pigs selected within each type for open field and novel object testing.

Type	Description	Rep 1	Rep 2	Total
Bold	Pigs quickly habituated quickly to saliva collection, and saliva was successfully collected on four collection days (as per 10)	8	4	12
Shy	Saliva not successfully collected in any, or less than half of collection days. The pig consistently ran away when approached during these sessions	8	4	12
Biter	Pig observed biting other pigs' tails or ears on at least three occasions (range 3–9), and more than any other pig in the pen, during weekly observation sessions over a 5 week period (as per 10)	8	4	12
Bitten	Either the tail or ear badly bitten (bleeding with open wound) a maximum of 1 week prior to testing	24	15	39

this was appropriate. The model included fixed effects of pig type ($n = 4$; BOLD, SHY, BITER, or BITTEN), treatment ($n = 2$; STRAW or TOY), sex ($n = 2$; male or female), and replicate ($n = 2$; 1 or 2). Pen and the day on which the test was carried out were included as random effects. To investigate the hypotheses that BOLD and BITER behaved differently to SHY and BITTEN pigs, a contrast statement was used. For this comparison, we hypothesized that the BOLD and BITER pigs would behave as a single cohort (labeled BRAVE) and the SHY and BITTEN pigs would behave differently to BRAVE, but similarly to each other, so were also considered a single cohort (labeled SCARED). Where data were not appropriate for analysis using linear models, the Kruskall-Wallis test was used instead (number of screams, attention toward the novel object, and time spent withdrawing from the object). The number of pigs performing these behaviors, or not, were also compared using Fishers exact test.

For analysis of salivary cortisol, a similar linear mixed model to that described above was used, with the addition of sampling time (before or after the behavior tests) as a repeated measure. The random effect of plate was also included in this analysis.

In all analysis using linear models, residuals were examined to verify normality and homogeneity of variances. Differences in least squares means were investigated using the t-test, using Tukeys adjustment for multiple comparisons. Degrees of freedom were estimated using Kenwood–Rogers adjustment. Differences were considered significant at $P \leq 0.05$. Tendencies toward significance ($0.05 < P \leq 0.10$) are also presented. Data are presented as Least Squares means ± SE.

RESULTS

Correlations

All correlations and their significance levels are shown in **Table 3**. In brief, exploration (explore floor, wall, or object in **Table 1**) was negatively correlated with walking, standing still, the number of locations entered, the duration spent at the door, and the number of screams. Grunts and screams were positively correlated. The number of locations traversed was positively correlated with grunts and screams, yet negatively correlated with the number of eliminations. The duration by the door was positively correlated with both grunts and screams. The number of grunts was negatively correlated with the latency to interact with the object, and positively correlated with the time spent interacting with it. The time directing attention toward the object was positively correlated with the latency to interact with it, and the time spent withdrawing from it, and negatively correlated with the time spent interacting with it. The time spent interacting with it was conversely negatively correlated with the time spent withdrawing and the latency to interact.

Principle Component Analysis

When the 11 behavior variables were included in an initial PCA, the first three components had an eigenvalue >1 and the scree plot indicated that there was a large separation between component three and four. These initial three components contributed 24.3, 18.2, and 15.0% of the variance in the dataset (combined: 57.4%). However, the variable "exploration" was found to load onto both component 1 and 3, and as such the analysis was re-run after removal of this variable. After the new analysis, the first three components still had an eigenvalue >1, there was a clear separation between component 2 and 3 according to the scree plot. These three components contributed 23.8, 18.6, and 12.5% of the variance in the dataset, totalling 54.5%. However, "walk" the number of locations visited, and the time spent withdrawing from the object now loaded onto two components, and as such were removed from a third analysis. This analysis showed that 3 components with an eigenvalue >1, yet no clear separation between any components. Components 1, 2, and 3 contributed 27.8, 20.1, and 14.7% respectively, totalling 62.6%), but only two variables loaded onto components 2 and 3. Moreover, the three remaining variables in the dataset which were contributed from the novel object test (attention toward the object, latency to touch the object, and time spent interacting with it) all loaded on to component 1. As such, the variables which contributed to the PCA did not meaningfully reduce the dataset into a smaller number of components which could be compared.

Open Field Test

There were no pair-wise differences between the pig type categories in any of the measurements taken during the OF test (**Table 4**). However, there tended to be an overall effect of pig type on the time spent exploring ($P = 0.09$) and there was also an effect on the time spent by the door ($P = 0.05$). For both of these, there was a difference between the pigs that were SHY or BITTEN pigs when compared to those that were BOLD or BITER pigs (**Figures 1A,B**). The BOLD and BITER pigs spent more time exploring ($P < 0.05$) and less time by the door ($P < 0.01$) than the SHY and BITTEN pigs

TABLE 3 | Correlations between the behaviors measured in the open field and novel object tests.

	Walking[a]	Exploration[a]	Standing[a]	No locations[a]	Duration by door[a]	Grunts[a]	Screams[b]	Elimination[a]	Attention[b]	Interaction[a]	Withdraw[b]	Latency to interact[b]
Walking[a]	**1**	**−0.668 <0.001**	0.003 0.98	**0.638 <0.001**	0.190 0.10	**0.283 =0.01**	**0.353 =0.001**	−0.175 0.13	−0.016 0.88	−0.044 0.71	−0.007 0.95	**−0.194 <0.05**
Exploration of floor, wall or object[a]		**1**	**−0.598 <0.001**	**−0.227 =0.05**	**−0.460 <0.001**	−0.194 0.10	**−0.347 <0.01**	0.162 0.17	0.118 0.32	0.101 0.39	0.103 0.38	0.075 0.41
Standing[a]			**1**	**−0.371 <0.001**	0.065 0.58	−0.155 0.18	−0.082 0.48	0.015 0.90	0.049 0.68	**−0.270 <0.05**	−0.025 0.83	**0.209 <0.05**
No. locations[a]				**1**	0.113 0.34	**0.278 <0.05**	**0.229 <0.05**	**−0.225 =0.05**	−0.047 0.69	0.008 0.95	−0.008 0.95	**-0.219 0.06**
Duration by door[a]					**1**	**0.233 0.042**	**0.375 <0.001**	−0.112 0.338	0.012 0.92	0.128 0.27	−0.156 0.18	−0.085 0.47
Grunts[a]						**1**	**0.340 <0.01**	0.054 0.65	−0.142 0.23	0.128 0.28	−0.185 0.11	**−0.272 <0.05**
Screams[b]							**1**	−0.062 0.60	−0.038 0.74	0.031 0.79	−0.131 0.27	−0.171 0.15
Elimination[a]								**1**	0.151 0.20	0.068 0.57	−0.051 0.67	−0.126 0.29
Attention[b]									**1**	**−0.235 <0.05**	**0.276 <0.05**	**0.328 <0.01**
Interaction[a]										**1**	**−0.283 <0.05**	**−0.596 <0.001**
Withdraw[b]											**1**	0.103 0.378
Latency to Interact[a]												**1**

[a]Correlations carried out using Pearsons correlation coefficient (r).
[b]Correlations carried out using Spearmans rank correlation coefficient (r_S).
Significance levels are indicated below correlation coefficients, and significant relationships highlighted in bold.

TABLE 4 | Results from the open field test and novel object test.

	BOLD	SHY	BITER	BITTEN	*P*-value
Open field test					
Walk	00:33 ± 00:11	00:39 ± 00:11	00:35 ± 00:11	00:42 ± 00:06	0.87
No. square transitions	23.12 ± 3.06	23.95 ± 3.06	23.95 ± 3.06	24.91 ± 1.71	0.96
Low grunts	38.22 ± 5.85	79.55 ± 5.85	33.63 ± 5.85	33.30 ± 3.37	0.63
Screams[a]	17% [0 (0–0)]	8% [0 (0–0)]	% [0 (0–0)]	23% [0 (0–0)]	0.38
Elimination	3.24 ± 0.47	1.74 ± 0.47	1.66 ± 0.47	2.40 ± 0.26	0.06
Novel object test					
Attention[a]	58% [2.8 (0.26–7.8)]	75% [1.34 (0–6.0)]	67% [2.2 (0–6.1)]	56% [2.1 (0–9.4)]	0.87
Latency to interact	00:51 ± 00:19	01:11 ± 00:19	01:12 ± 00:19	00:48 ± 00:10	0.55
Interaction duration	01:46 ± 00:17	00:57 ± 00:17	01:12 ± 00:17	01:11 ± 00:09	0.19
Withdraw[a]	33% [0 (0–8.7)]	42% [0 (0–1.76)]	17% [0 (0–0)]	10% [0 (0–0)]	0.10

Unless indicated otherwise, data are presented as Least Squares means ± standard error.
[a]Data presented as the percentage of pigs which performed the behavior, as well as the median and interquartile range of the duration for which it was performed, including 0 values. P-values are in relation to the Kruskall-Wallis test.

(**Figures 1A,B**). The SHY and BITTEN pigs also tended to spend more time standing still than the BITER and BOLD ($P = 0.1$; **Figure 1C**).

Novel Object Test

There was no effect of pig type on any of the measurements recorded during the NO test (**Table 4**).

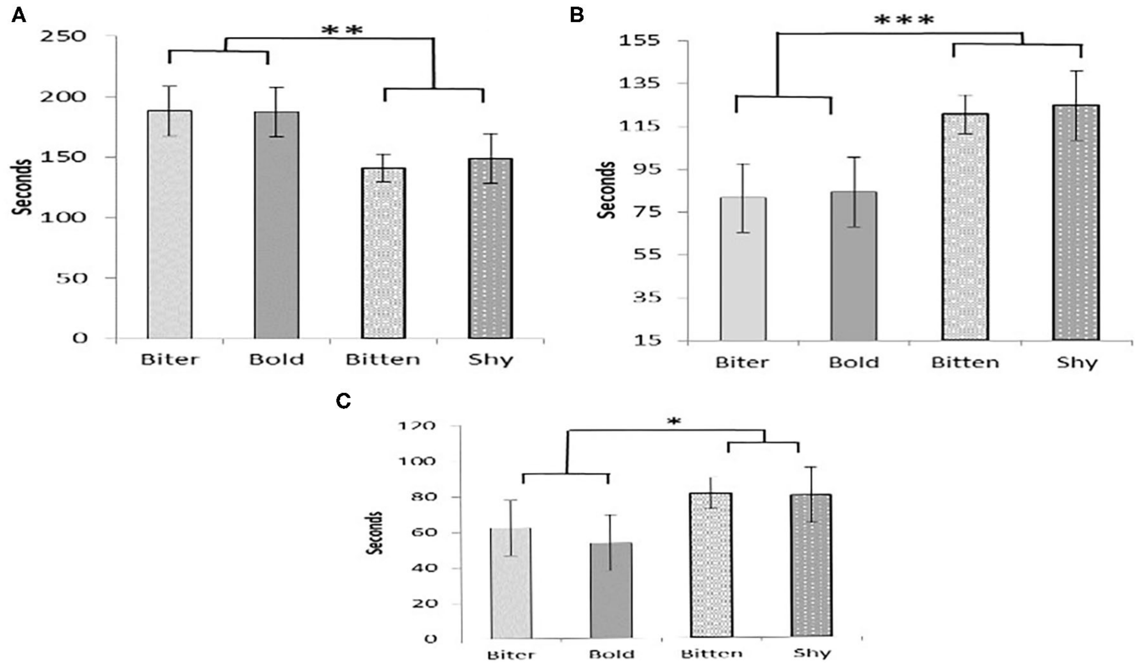


FIGURE 1 | The amount of time pigs in each category spent exploring **(A)**, by the door **(B)**, and standing still **(C)**. Tendencies and significant differences between BITER and BOLD with BITTEN and SHY pigs are indicated by * = $0.5 < P \leq 0.1$; ** = $0.01 < P \leq 0.05$, *** = $0.001 < P \leq 0.01$.

Salivary Cortisol

Saliva samples collected after the behavioral tests had higher cortisol levels than prior to the tests (0.289 ± 0.066 vs. 0.776 ± 0.068 μg/dl; $P < 0.001$). There was no effect of pig type on cortisol level, or interaction between pig type and whether the sample was taken before or after the behavior tests (**Figure 2**). However, the increase in salivary cortisol level was significant for SHY ($P < 0.05$) and BITTEN ($P < 0.001$) pigs, but only tended to increase for BITER and BOLD ($P = 0.08$ for both).

DISCUSSION

The aim of this work was to investigate firstly whether pigs which were fearful of human contact in their home pen (SHY) would display more fear related behavior in an open field and novel object test than pigs which were not fearful of human contact (BOLD) in the home pen. This was so that we could determine whether the expression of fear related behavior was consistent across scenarios. These data also provided a robust baseline with which to compare our "control" pigs, to bitten and biter pigs, as in previous studies "control" pigs were selected only on the basis that they were neither bitten nor biters. If behavior in the test situation aligned with pig category as determined in the home pen, then we can hypothesize that at least in the fear eliciting situations during which we observed pig behavior (forced human contact and OA and NO tests) these pigs behaved consistently. We then investigated whether pigs that had experienced tail or ear biting, would display the same fear related behaviors as human shy pigs, and whether biter pigs would behave in a similar fashion to human bold pigs. Thus, if the bitten pigs also respond in a similar fashion to SHY pigs in the behavior arena, and the biters in a similar fashion to BOLD, we can hypothesize that these pigs may also be in general more or less fearful across a range of situations. Overall, we found that our hypotheses were partly supported; for some, but not all, measures, pigs which were human shy, and pigs which were bitten, had responses similar to each other, and different to the responses of human bold and biters pigs, and primarily in the OF test. While we are aware of the limitations that our small sample size has created, we believe that this study helps to lay the foundations for the development of further work using a larger sample size, and particularly of biter pigs. Based on our experience with the current study, we also suggest that improvements such as more detailed analysis of vocalizations and the amount of exploration performed in the home pen could add value when selecting pigs for testing. This may help to identify more distinct differences between pig types, and ultimately to determine whether based on their innate personality traits, whether a pig is predisposed to be a biter or a victim.

In agreement with our hypothesis, pigs which we expected would be less fearful in the behavior tests (BOLD and BITER) spent more time exploring the arena, less time by the door, and less time standing still, than other pig types (SHY and BITTEN). Indeed time spent exploring the arena was negatively correlated with both standing still, and time spent by the door. Exploration is considered a normal and highly motivated behavior for pigs, as under natural conditions they spend a high proportion of their time performing exploratory and rooting behavior in their search

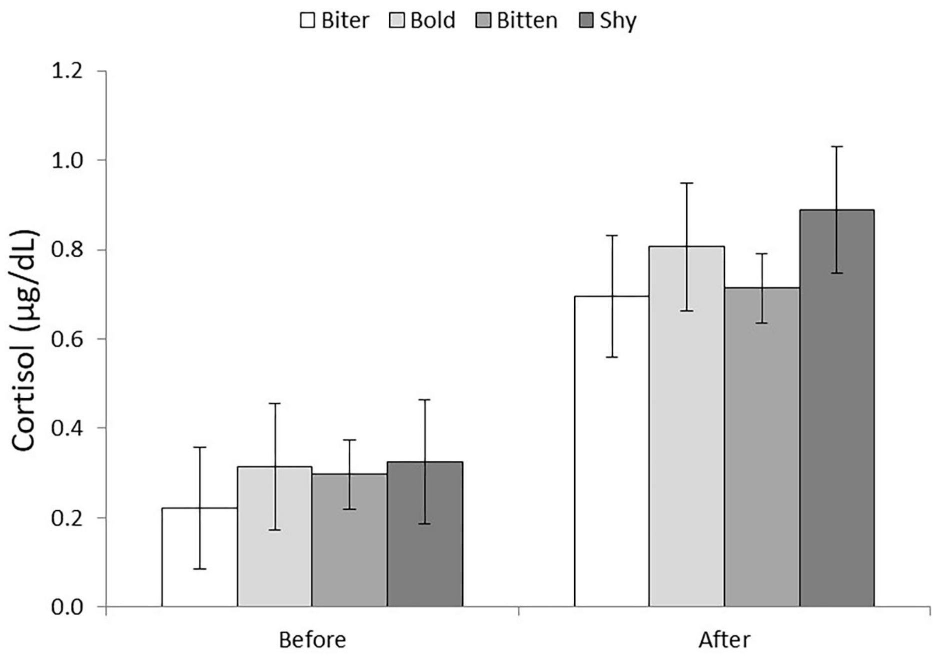


FIGURE 2 | Salivary cortisol levels (µg/dl) of the four pig types (BITER, BOLD, BITTEN, SHY) before and after the open field and novel object tests.

for food (15). Thus, higher level of performance of this behavior in the human bold, and biter pigs, indicates that their behavior was less inhibited by the potentially fear inducing experience of the open field test than the pigs which were human shy and bitten.

Biter pigs have previously shown to display more exploration behavior in a home pen environment; Ursinus et al. (12) found that time spent exploring pen-mates or the environment was associated with being classified as a tail biter, for finisher pigs which were managed in housing conditions similar to those in our study. Similarly, Zonderland et al. (16) observing that biter pigs interacted with enrichment more than other pen-mates during the 6 days prior to a tail biting outbreak. The pigs in the current study were categorized as being "biters" if they were regularly (>3 times) observed biting over 5 weeks of observation, even though there were no tail biting outbreaks in the pens, and as such they may have simply been fulfilling a high motivation to explore, rather than engaging in compulsive or highly damaging behavior. In contrast, it is possible that bitten pigs may have been less likely to engage in exploration if they were experiencing pain, stress, or sickness behavior or were avoiding biters. Human shy pigs appeared to be consistent in their response to a non-usual situation (a human attempting interaction, and the behavior tests) in that their responses are commonly interpreted as being indicative of fear. In future studies, it would be useful to consider variations in exploration behavior amongst BITER and BITTEN pigs in the home pen prior to the tests.

The lack of the ability of the principle component analysis to reduce the behavior data to meaningful components, as was reported in previous studies (e.g., 4), was disappointing. At the same time, there were yet several correlations between variables, which when examined, can provide some level of insight into the relationship between the measurements. The fact that exploration was negatively correlated with walking, standing still, the number of locations entered and the duration spent at the door is somewhat obvious, as these behaviors are in the main mutually exclusive. Its negative association with the number of screams is also somewhat intuitive, as screams can indicate stressful situations (17). In contrast, grunts are generally considered to be indicative of a "positive" vocalization in pigs.

Nevertheless, in the current study grunts and screams were positively correlated, and both were positively correlated with duration by the door, a behavior which we hypothesize to be related to fear, as well as time spent walking and the number of locations visited in the open field test. These data could indicate that pigs which traveled through the arena without taking time to explore, were attempting to find an escape route. Indeed, Mormede et al. (18) considered that in the novel environment test, fearfulness is usually observed either as low activity (behavioral inhibition) or high activity (behavioral activation, usually associated with escape attempts) together with many squeals. Thus, in this instance, grunting may not represent a positive experience. Marchant et al. (19) found that single grunts could be sub-divided into two types based on sound amplitude profile, with short single grunts being associated with investigatory behavior and on-going single grunts potentially being a form of contact call. Detailed examination of vocalizations was not possible during this study; it may have been more insightful if vocalizations were recorded, and submitted through software which could distinguish the precise components of the sound, as in the studies by Marchant et al. (19) and Schon et al. (20).

Surprisingly, given the associations from the open field test, grunts were associated with a low latency to interact with the novel object, and a longer time interacting with it. Thus, these vocalizations appeared to be more associated with a lack of fear in relation to exposure to a novel object. It is important to remember that the open field and novel object tests may measure different aspects of fear or anxiety. Moreover, grunts and screams were recorded during the former test, and not during the latter. By the time the novel object test occurred, it is possible that the pigs had habituated to the test arena. Moreover, the detailed comparison of fear tests carried out by Andersen et al. (21) identified that aversion to novelty and measures of activity were two separate components. Vocalizations however, were not recorded. Thus, it is possible that vocalizations may or may not have been associated with one or both of these components; in our study, vocalizations were associated with both increased locomotion, and no aversion to novelty.

The tests we used were unfamiliar to all pigs, with no pigs having been previously isolated from their conspecifics. It was therefore unsurprising that cortisol levels were higher after the tests than prior. Although, shy and bitten pigs did not show higher levels than biter or bold pigs either before or after the test, these were the only two categories which had significantly higher levels after the test than before. This confirms our hypothesis that these animals would have a greater stress response to the tests, or indeed the taking of a cortisol sample (which in itself could prove to be fear or stress inducing), than the human bold or biter pigs.

At the time of testing, no major tail biting outbreaks had occurred [e.g., 21.4–25% of pigs per pen with fresh dripping blood or a tail damage score of 3- (22)], and as such the biter pigs may not have been the type of "compulsive" biter that instigates an outbreak, but rather pigs which engage in excessive levels of two stage biting, or sudden forceful biting when aiming to access a resource. Further work, with a larger sample size of "biter" pigs, using pigs which were not tail docked, and using more detailed behavior observations to categorize pig types, would be a worthy exercise.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The animal study was reviewed and approved by Teagasc Animal Ethics Committee (TAEC89/2015). Written informed consent was obtained from the owners for the participation of their animals in this study.

AUTHOR CONTRIBUTIONS

AH and KO'D contributed to the conception, design of the study, and completed the writing of the manuscript. AH, J-YC, and KO'D conducted all aspects of the experimental testing. KO'D organized the database and completed all the statistical analysis. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This project was funded under the Department of Agriculture, Food and Marine's Competitive research programmes.

ACKNOWLEDGMENTS

We would like to acknowledge Laura Boyle, Oceane Schmitt, Dave Clarke, and Oliver Clear for all their help with the experimental testing and the staff and owner of the commercial farm were the experiment took place. We would also like to thank the two reviewers for their constructive comments which greatly improved the manuscript.

REFERENCES

1. Schroder-Petersen DL, Simonsen H. Tail biting in pigs. *Vet J*. (2001) 162:196–210. doi: 10.1053/tvjl.2001.0605
2. Harley S, Boyle LA, O'Connell NE, More SJ, Teixeira DL, Hanlon A. Docking the value of pigmeat? Prevalence and financial implications of welfare lesions in Irish slaughter pigs. *Anim Welf*. (2014) 23:275–85. doi: 10.7120/09627286.23.3.275
3. Taylor NR, Main DC, Mendl M, Edwards SA. Tail-biting: a new perspective. *Vet J*. (2010) 186:137–47. doi: 10.1016/j.tvjl.2009.08.028
4. Ursinus WW, Van Reenen CG, Kemp B, Bolhuis JE. Tail biting and tail damage in pigs and the relationship with general behaviour: predicting the inevitable? *Appl Anim Behav Sci*. (2014)156:22–36. doi: 10.1016/j.applanim.2014.04.001
5. Zupan M, Zanella AJ. Peripheral regulation of stress and fear responses in pigs from tail-biting pens. *R Braz Zootec*. (2017) 46:22–8. doi: 10.1590/s1806-92902017000100006
6. Zupan M, Janczak A, Framstad T, Zanella AJ. The effect of biting tails and having tails bitten in pigs. *Physiol Behav*. (2012) 106:638–44. doi: 10.1016/j.physbeh.2012.04.025
7. Forkman B, Boissy A, Meunier-Salaun M-C, Canali E, Jones RB. A critical review of fear tests used on cattle, pigs, sheep, poultry and horses. *Physiol Behav*. (2007) 92:340–74. doi: 10.1016/j.physbeh.2007.03.016
8. Mieloch FJ, Nietfeld S, Straßburg C, Krieter J, Grosse Beilage E, Czycholl I. Factors of potential influence on different behavioural tests in fattening pigs. *Appl Anim Behav Sci*. (2020) 222:104900. doi: 10.1016/j.applanim.2019.104900
9. O'Malley CI, Turner SP, D'Eath RB, Steibela JP, Bates RO, Ernst CW, et al. Animal personality in the management and welfare of pigs. *Appl Anim Behav Sci*. (2018) 218:104821. doi: 10.1016/j.applanim.2019.06.002
10. Haigh A, Yun-Chou J, O'Driscoll K. An investigation into the effectiveness of compressed straw blocks in reducing abnormal behaviour in growing pigs. *Animal*. (2019) 13:2576–85. doi: 10.1017/S1751731119000715
11. Valros A, Palander P, Heinonen M, Munsterhjelm C, Brunberg E, Keeling L, et al. Evidence for a link between tail biting and central monoamine metabolism in pigs (*Sus scrofa domestica*). *Physiol Behav*. (2015) 143:151–7. doi: 10.1016/j.physbeh.2015.02.049
12. Ursinus WW, Van Reenen CG, Reimert I, Bolhuis JE. Tail biting in pigs: blood serotonin and fearfulness as pieces of the puzzle? *PLoS ONE*. (2014) 9:e107040. doi: 10.1371/journal.pone.0107040

13. Hunter EJ, Jones TA, Guise HJ, Penny RHC, Hoste S. Tail biting in pigs 1: the prevalence at six UK abattoirs and the relationship of tail biting with docking, sex and other carcass damage. *Pig J*. (1999).
14. O'Rourke N, Hatcher L. *A Step-by-Step Approach to Using SASR® for Factoral Analysis and Structural Equation Modelling*. 2nd Ed. Cary, NC: SAS institute (2013).
15. Stolba A, Wood-Gush DGM. The behaviour of pigs in a semi-natural environment. *Anim Prod*. (1989) 48:419–25. doi: 10.1017/S0003356100040411
16. Zonderland JJ, Schepers F, Bracke MBM, den Hartog LA. Characteristics of biter and victim piglets apparent before a tail-biting outbreak. *Animal.* (2011) 5:767–75. doi: 10.1017/S1751731110002326
17. Vandermeulen J, Bahr C, Tullo E, Fontana I, Ott S, Kashiha M, et al. Discerning pig screams in production environments. *PLoS ONE.* (2015) 10:e0123111. doi: 10.1371/journal.pone.0123111
18. Mormède P, Dantzer R, Bluthe RM, Caritez JC. Differences in adaptive abilities of three breeds of Chinese pigs. behavioural and neuroendocrine studies. *Genet Sel Evol.* (1984) 16:85. doi: 10.1186/1297-9686-16-1-85
19. Marchant JN, Whittaker X, Broom DM. Vocalisations of the adult female domestic pig during a standard human approach test and their relationships with behavioural and heart rate measures. *Appl Anim Behav Sci.* (2001) 72:23–39. doi: 10.1016/S0168-1591(00)00190-8
20. Schon PC, Puppe B, Manteuffel G. Automated recording of stress vocalisations as a tool to document impaired welfare in pigs. *Anim Welf.* (2004) 13:105–10.
21. Andersen IL, Faerevik G, Boe KI, Janczak AM, Bakken M. Effects of diazepam on the behaviour of weaned pigs in three putative models of anxiety. *Appl Anim Behav Sci*. (2000) 68:121–30. doi: 10.1016/S0168-1591(00)00098-8
22. Chou JY, O'Driscoll K, D'Eath RB, Sandercock DA, Camerlink I. Multi-step tail biting outbreak intervention protocols for pigs housed on slatted floors. *Animals.* (2019) 9:582. doi: 10.3390/ani9080582

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.

frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 24 September 2020
doi: 10.3389/fvets.2020.575795

Development of an Early Warning System for Owners Using a Validated Health-Related Quality of Life (HRQL) Instrument for Companion Animals and Its Use in a Large Cohort of Dogs

*Vinny Davies[1], E. Marian Scott[2], M. Lesley Wiseman-Orr[2,3], Andrea K. Wright[4] and Jacqueline Reid[5,6]**

[1] School of Computing Science, University of Glasgow, Glasgow, United Kingdom, [2] School of Mathematics and Statistics, University of Glasgow, Glasgow, United Kingdom, [3] School of Education, University of Glasgow, Glasgow, United Kingdom, [4] Outcomes Research, International Centre of Excellence, Zoetis, Dublin, Ireland, [5] NewMetrica Ltd., Glasgow, United Kingdom, [6] School of Veterinary Medicine, University of Glasgow, Glasgow, United Kingdom

OPEN ACCESS

Edited by:
Edward Narayan, The University of Queensland, Australia

Reviewed by:
Jo Hockenhull, University of Bristol, United Kingdom
Jenna Kiddie, Dogs Trust, United Kingdom

***Correspondence:**
Jacqueline Reid
Jacky.reid@newmetrica.com

Specialty section:
This article was submitted to Animal Behavior and Welfare, a section of the journal Frontiers in Veterinary Science

Received: 24 June 2020
Accepted: 12 August 2020
Published: 24 September 2020

Citation:
Davies V, Scott EM, Wiseman-Orr ML, Wright AK and Reid J (2020) Development of an Early Warning System for Owners Using a Validated Health-Related Quality of Life (HRQL) Instrument for Companion Animals and Its Use in a Large Cohort of Dogs. Front. Vet. Sci. 7:575795. doi: 10.3389/fvets.2020.575795

Preventive measures in human healthcare are recognized as a means of providing early detection of disease, however, the veterinary profession has not been as effective in communicating the benefits of preventive measures to pet owners. Readily available pet healthcare information on the internet, owners not understanding that regular health evaluations can ensure the well-being of their pets and owners confusing the signs of chronic disease with normal aging have contributed to declining numbers of veterinary visits. The use of web-based generic health–related quality of life (HRQL) measures to evaluate health status (wellness) remotely could facilitate veterinary preventive medicine. This publication describes the development and practical application of an integrated alert system for an online generic HRQL measurement instrument (VetMetrica™) which generates scores in four domains of HRQL—Energetic/Enthusiastic (E/E), Happy/Content (H/C), Active/Comfortable (A/C), and Calm/Relaxed (C/R)—for 2 age groups (young/middle-aged, ≤7 years and old, ≥8 years). The alert provides an early warning, via email to owners, that a potentially significant deterioration in health status has occurred. The model accurately predicted the health status of 93 and 83% of sick young/middle aged and old dogs respectively, with healthy dogs predicted with 83% accuracy. HRQL data, collected via a white-labeled veterinary clinic branded app designed to facilitate connected care between owner and veterinarian, were analyzed for 6,108 dogs, aged between 6 weeks and 16 years. Of these 5,002 were deemed to be in perfect health by their owners, yet the alert was triggered for 1,343 (27%) of these, 75% of which were young/middle-aged and 25% were old, indicating that acute injuries notwithstanding, many middle aged dogs may have been suffering from undetected chronic disease such as osteoarthritis. This work has demonstrated that the use of VetMetrica™ delivered via the PetDialog™ app, which supports 24/7 remote health monitoring is an efficient way for vets to provide all their owners with the opportunity to monitor their animal's wellness throughout their lifetime, providing the vet with a mechanism to identify health problems early while stimulating owners to be more proactive in seeking veterinary attention.

Keywords: Health-Related Quality of Life (HRQL), owner questionnaire, smartphone app, dogs, early warning system, wellness, preventive medicine

INTRODUCTION

Preventive medicine in the food animal sector plays an important role in preventing and controlling important diseases of food animals and humans and involves a number of disciplines including epidemiology and public health. Similarly, in human healthcare preventive measures have long been recognized as a means of providing early detection of disease or avoiding it completely. In that regard, recent research developments highlight the value of Patient–Generated Health Information (PGHI), described as data created and recorded, by patients, family or personal caregivers, whereby health data can be shared with a health care provider via a smartphone health app or a patient portal accessed in the home (1). In human healthcare, patient symptoms, health status, and quality of life are increasingly monitored in this way. Receiving information directly from the patient or caregiver can be valuable in many ways. It can strengthen the patient/caregiver-clinician relationship, providing opportunities for shared decision making as well as capturing data that otherwise would be missed by traditional means. However, veterinarians working with companion animals have not been as effective in communicating these benefits of preventive measures to pet owners and this has contributed to a decade long trend toward declining numbers of pet owner visits to veterinarians. Even though the number of cats and dogs has increased and continues to do so, the footfall in veterinary practice has declined. In the US alone the cat and dog population increased >36% in the 10 years prior to 2006, yet American Animal Hospital Association data show that between 2001 and 2009 there was a decrease of 17% in the median number of active clients seen by full-time veterinarians (2). According to the 2011 Bayer Veterinary Care Usage Study this decline can, in part, be attributed to the readily available pet healthcare information on the internet, which for many owners is their first port of call when their pets show signs of illness. At best, this results in a delayed veterinary visit when symptoms persist, but may represent a missed visit if the animal improves. Perhaps more worrying is the fact that another major reason for the decline is lack of understanding on the part of pet owners that regular health evaluations will help ensure the well-being of their pets. Unfortunately, many pet owners still associate veterinary care primarily with vaccinations and the treatment of acute diseases or injuries, tending to confuse the signs of chronic illness with the aging process. Furthermore, with the increase in the companion animal population has come a disproportionate increase in the number of geriatric animals and a consequent increase in the incidence of chronic disease. Indeed according to the Banfield Pet Hospital, state of pet health reports (3–5), there has been a substantial increase in the incidence of diabetes and obesity, which along with increase dental disease, parasitic infestation, and behavioral issues can be avoided with routine preventive care (6). Moreover, early identification and treatment of many diseases can reduce the need for costly interventions later. Clearly, there needs to be an increase in communication between veterinary surgeons and their clients about the advantages of preventive care and this has been recognized by the American Animal Hospital Association and American Veterinary Medical Association who introduced their Preventive Healthcare Guidelines in 2011. Similarly, in 2010 the North American Veterinary Medical Education Consortium recommended that there should be more focus on wellness and disease prevention in veterinary undergraduate education. These measures are very valuable, but Spofford et al. (7) suggest that there is still a need for research "to determine the impact of preventive health care in animals and to distinguish effective preventive health-care services from less effective and ineffective ones." These authors suggest that there is a need for health—related quality of life (HRQL) evaluations that capture useful information about health that is not readily obtained from the medical case record and that, in particular, generic HRQL instruments that assess wellness could be used to measure the impact of preventive health-care services. In contrast to disease–specific HRQL instruments which have an application limited to sick populations, generic instruments measure the health status of healthy as well as sick subjects and are the only option when co-morbidities exist in the same subject.

Previously, we have reported the development, validation, and reliability of owner–reported generic HRQL instruments (VetMetrica™) for the dog (8, 9) and the cat (10), both of which were generated using data from owners of healthy and sick animals. These are structured questionnaires containing simple behavioral questions for completion online by the owner, with scores generated in four domains of quality of life for the dog (Energetic/Enthusiastic, Happy/Content, Active/Comfortable, and Calm/Relaxed) and three for the cat (Vitality, Comfort, Emotional Wellbeing). In QOL measurement there are two kinds of variables that can be measured—indicator and causal (11). Causal variables like "vomiting" impact the QOL, but don't tell us anything about it, whereas indicator variables like "energy" don't affect it, but do give us information about it. In general, other HRQL tools measure the physical limitations imposed by disease, whereas the VetMetrica™ questionnaires consist only of indicator variables that measure the emotional component of QOL—how the animal "feels." In this regard, they are unique. The web–based system is compatible with all mobile platforms and since 2014 has been made available to pet owners through the smartphone app PetDialog™ (compatible with iOS and Android devices), developed by Zoetis. PetDialog™ is a white-labeled veterinary clinic branded app designed to facilitate connected care between an owner and a veterinarian. The veterinarian can view the responses from their own clients in a real-time dashboard accessed through a web-platform called VetSupport+. The owner receives timely push-notification alerts to answer HRQL questions within the app where they can also track the results over time. Additionally, the owner, via the app, or the veterinarian, via VetSupport+, can set up medication reminders to assist with compliance of medical recommendations. The PetDialog™ app and Vet Support+ make up the ecosystem that supports 24/7 remote health monitoring of pets.

A key property of any health-related measure is its interpretability, because, without that, the instrument is of little or no practical use when it is used to measure the impact of healthcare interventions. In the medical field interpretability of health measures is an important focus of current research (12), and although there is no consensus amongst our medical colleagues regarding the optimum method of determining interpretability, a number of methods have been described, including relating the HRQL scores to those of a specific population, facilitating judgment as to whether an observed score is typical of what would be expected for that population (norm-based scoring). Norm-based comparisons can be related to the general population, to sub-populations with shared demographics such as age or gender or to a population with a disease (13). Initially, the outputs for VetMetrica™ were reported as "raw" 0–6 scores for each domain, but more recently the calculation of norm-based scores has been reported for the dog tool to improve its interpretability (14). Domain scores are normed to the age—related average healthy dog where the classification is young/middle aged (≤7 years) and old (≥8 years).

While those features which support interpretability of an instrument are fundamentally important, its functionality is markedly enhanced if there is an associated "call to action" which will alert the patient/carer/clinician to the fact that there has been a change in the state of health. Currently in companion animal medicine there is a focus on measuring health status using advanced wearable monitors that can detect changes in activity and a variety of other activities such as drinking, scratching, sleeping, etc. and depending on the monitor will alert the owner to a change in these parameters. These objective measures of physical functions and changes therein may be useful under certain circumstances, especially where changes in mobility or scratching are important indicators of orthopedic or skin disease, and a change in drinking may indicate renal disease or diabetes. However, there are several commonly occurring diseases in animals, such as cancer and obesity that do not exhibit the acute changes in these physical manifestations that will be highlighted with an activity monitor. In these cases, a generic HRQL instrument with an inbuilt health alert would provide useful information for clinician and owner alike.

The aims of this study were firstly to develop an algorithm which will indicate whether a dog is healthy or unhealthy (Phase 1) and incorporate that in the VetMetrica™ dog software to provide a 'call to action' to the owner to consult their vet (Phase 2), finally reporting on its use in a large cohort of dogs using the PetDialog™ App (Phase 3).

PHASE 1: DEVELOPMENT OF THE ALGORITHM USING RAW DOMAIN SCORES

Materials and Methods

All data were retrospective having been collected in a variety of previous studies. All studies were approved by the Ethics Committee of the University of Glasgow Veterinary School and owners gave informed consent for participation in these studies. Dogs (170 ≤ 7 years and 252 ≥ 8 years) were recruited from the University of Glasgow Small Animal Hospital and private veterinary practices. Data used to develop the alert algorithm (**Supplementary Information**). The health status of all dogs was assessed by the attending veterinary surgeon and the only inclusion criterion for sick dogs was that they were suffering from a chronic condition likely to affect their QOL. There were no exclusion criteria. Owners received an information sheet with the following wording: "*NewMetrica is a small company run by a vet that develops questionnaires to enable us to determine how animals are feeling. We do this by asking you, the owner, certain questions about your dog's behavior. We have developed a short questionnaire called VETMETRICA for this purpose which can be completed online in around five minutes in your own home. We need a large number of owners to complete it so that we can analyze all the information to allow us to test how well the questionnaire works. We would be very grateful for your help with this.*"

Owners completed the Vetmetrica™ HRQL instrument for dogs which contained 22 items, each of which comprised a descriptor (e.g., "active") with a 7-point Likert rating scale, 0–6 (with 0 meaning "not at all" and 6 meaning "could not be more"), which were used to determine a raw HRQL score (0–6) in each of four domains (Energetic/Enthusiastic, Happy/Content, Active/Comfortable, and Calm/Relaxed) (9, 14). The frequency of assessments for each dog was set by the vet with a minimum period of 2 weeks, however, analysis was restricted to the first assessment for each dog.

Multivariable logistic regression (15) was used to model the relationship between the four HRQL domain scores as covariates or independent variables and the corresponding binary health status of the dog as the response or dependent variable. This was done separately for young/middle aged (≤7 years) and old (≥8 years) dogs (41 sick young/middle aged dogs, 222 sick old dogs, respectively). The statistical model was fit in R (https://www.R-project.org/) using the glmnet package (16). Variable selection was carried out using 10-fold Cross Validation to identify the HRQL domains to be used in each of the models, based on how well they predict the health status of the dogs (17). The performance of the models in predicting the health status of a dog can be evaluated by looking at the sensitivity and specificity. To assess the sensitivity and specificity we must first identify a cut-off value in the regression equation above/below which a dog would be classified as healthy/sick. Varying the cut-off means that different numbers of dogs get a regression score of <0 (which are then classified as sick and the Contact Vet Flag is raised) or more than zero (which are then classified as healthy and no Contact Vet Flag is raised). Sensitivity measures the proportion of positives, i.e., the proportion of sick dogs correctly classified as sick. Specificity measures the proportion of negatives, i.e., the proportion of healthy dogs correctly classified as healthy. The higher the sensitivity and specificity the better however the two measures are dependent and so must be considered jointly. Receiver Operating Characteristic (ROC) scatterplots provide a graphical illustration of the performance over the different cut-offs [by plotting and connecting sensitivity vs. (1-Specificity) values for each cut-off]. The optimum cut-off is the point which is closest to the top left corner of the scatterplot since this represents the point on the ROC curve closest to optimal performance, i.e., maximum sensitivity and specificity).

Results

The final Vet Alert models for dogs ≤7 years (young/middle aged) and those ≥8 years (old) both contained the Energetic/Enthusiastic and Active/Comfortable HRQL domain variables, with the model for the young dogs also containing the Happy/Content domain. The results for the Vet Alert models in terms of sensitivity, specificity, and classification accuracy for the old and young/middle aged dogs are shown in **Table 1**. **Figure 1** shows the ROC curves for each of the models, with the sensitivities and specificities from the best cut-off identified by the large X and corresponding to the results in **Table 1**.

TABLE 1 | Sensitivities, specificities, and classification accuracy for the health alert models for old and young/middle-aged dogs.

Age	Sensitivity	Specificity	Classification Accuracy
Old	0.93	0.83	0.92
Young/Middle-aged	0.83	0.83	0.83

PHASE 2: INTEGRATION OF THE ALGORITHM, WITH A CASE EXAMPLE INCLUDED FOR ILLUSTRATIVE PURPOSES

Materials and Methods

The algorithm was embedded in the VetMetrica software by software developers Kyria Ltd. (http://kyria.co.uk/) to create a new output as "Contact Vet Score." If this score is positive, then no alert is raised (Contact Vet Flag—False). If it is negative then the Contact Vet Flag is raised (Contact Vet Flag—True) and in this case, on completion of the assessment, the owner receives a message to contact their vet.

Results

Figure 2A shows a section of the longitudinal results from an 11 year old dog which was suffering from well-controlled inflammatory bowel disease (IBD), as they are reported to the vet (normalized scores only). All four domains are individually color coded and normalized HRQL scores are shown on the y axis, where 50 represents the age-related average healthy dog. The dotted line at 44.8 represents the threshold above which 70% of healthy dogs in the appropriate age group will score. HRQL scores were very consistent until mid June when there was a decrease in Happy/Content and again until early to mid-August when all four domain scores showed a considerable decrease, recovering by early September. Because the Vet Flag is raised based on the values of the Energetic/Enthusiastic and Active/Comfortable domains only in dogs ≥8 years, the initial decrease in Happy/Content did not raise the flag.

Figure 2B shows the corresponding excerpt from the VetMetrica data extract, with assessment date, domain scores, Contact Vet Score, and Contact Vet Flag. The green highlight refers to the initial decrease in Happy/Content which occurred on

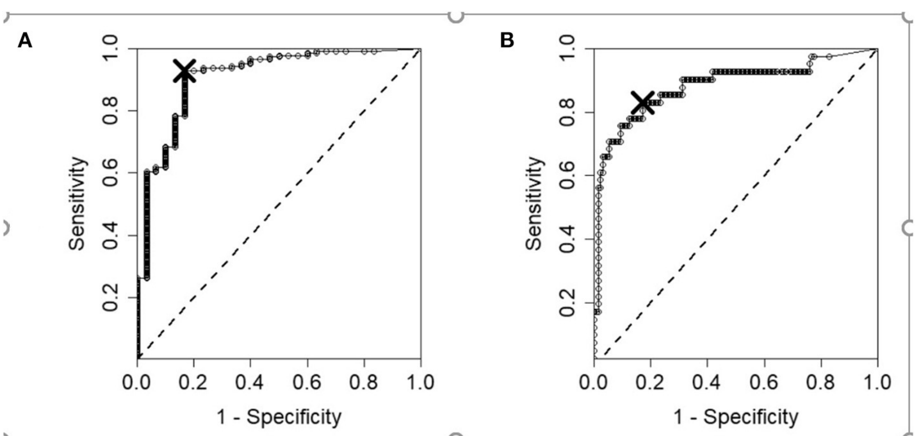
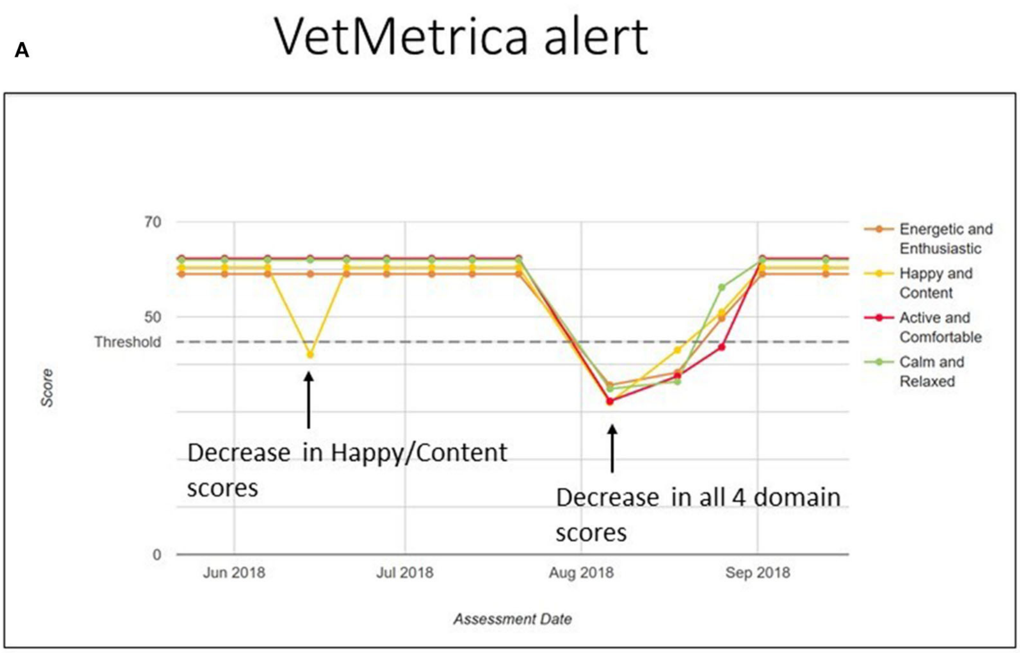


FIGURE 1 | ROC curves showing the results of the health alert model for the **(A)** old and **(B)** young/middle-aged dogs. The cut-offs chosen by the top left corner method are shown as large crosses.



B

1	Date completed	Energetic/Enthusiastic	Happy/Content	Active/Comfortable	Calm/Relaxed	Contact vet score	Contact vet flag
2	06/06/2018	59.0676	60.4527	62.3168	62.0292	0.739	FALSE
3	14/06/2018	59.0676	42.0851	62.3168	62.0292	0.739	FALSE
4	20/06/2018	59.0676	60.4527	62.3168	62.0292	0.739	FALSE
5							
6							
7	05/08/2018	35.6867	32.0128	32.3044	34.9134	-1.057	TRUE
8	17/08/2018	38.3454	43.0575	37.5474	36.4359	-0.6064	TRUE
9	25/08/2018	49.7204	51.0198	43.6836	56.2482	0.0936	FALSE
10	01/09/2018	59.0676	60.4527	62.3168	62.0292	0.739	FALSE

FIGURE 2 | (A) A section of the longitudinal results from an 11 year old dog which was suffering from well controlled inflammatory bowel disease (IBD), as they are reported to the vet (normalized scores only). All four domains are individually color coded and normalized HRQL scores are shown on the *y*-axis, where 50 represents the age-related average healthy dog. The dotted line at 44.8 represents the threshold above which 70% of healthy dogs in the appropriate age group will score. **(B)** A the corresponding excerpt from the VetMetrica data extract, with assessment date, domain scores, Contact Vet Score, and Contact Vet Flag. The green highlight refers to the initial decrease in Happy/Content which occurred on the 14th June and the yellow highlight shows the two occasions when the Vet Flag was raised, and an alert was issued to the owner.

the 14th June and the yellow highlight shows the two occasions when the Vet Flag was raised, and an alert was issued to the owner.

Figure 3 is a screenshot of the final "thank you" page of the assessment where the owner is alerted to the fact that they should seek advice from their vet.

PHASE 3 COLLECTION OF DATA AND REPORTING OF NORMALIZED SCORES VIA AN APP

Materials and Methods

Health-related quality of life data were collected from dog owners using VetMetrica™ for dogs via an app (PetDialog™, Zoetis). The VetMetrica™ dog HRQL instrument was incorporated as one of several features in the PetDialog™ app, which also required owners to input their dog's date of birth (DOB), breed, sex, and neutering status. Engagement with the HRQL instrument was entirely at the discretion of owners and uptake was not assessed. The PetDialog™ app was made available to pet owners in the United Kingdom and The Netherlands via veterinary practices and was only accessible using a unique practice code. Due to data protection regulations, the data were anonymized such that the demographic details and geographical location of each dog and owners were unknown. Normalized scores in Energetic/Enthusiastic, Happy/Content, Active/Comfortable, and Calm/Relaxed (14) were reported and the owner's impression of their dog's health status was determined by asking an additional question "Is your dog in perfect health—yes or no" after the 22 questions comprising the assessment were submitted. A definition for perfect health was

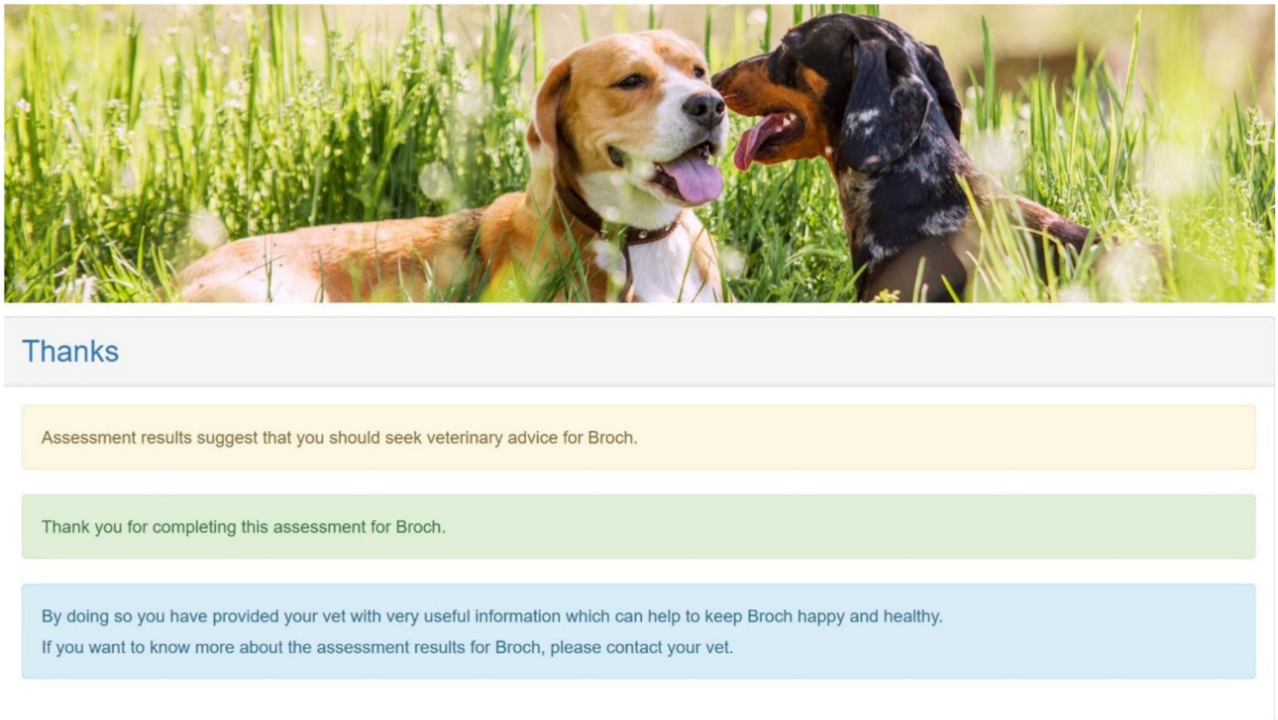


FIGURE 3 | A screenshot of the final "thank you" page of the assessment where the owner is alerted to the fact that they should seek advice from their vet.

not included. This question was posed for research purposes only and did not form part of the HRQL assessment. Analysis was restricted to the first assessment for each dog.

Results

Health-related quality of life data were collected over a period of 72 months (2014–2020) for 6,286 dogs. Data collected via the PetDialog App (**Supplementary Information**) Data were removed for 178 dogs for which DOB was deemed unreliable. For the remaining 6,018 dogs, there were 2,889 females of which 344 were neutered and 3,219 males of which 354 were neutered. Sixty-eight dog breeds were represented, and these were classified by the developers of the VetMetrica^{TM} system as follows, small (7%), medium 30%), large (33%), and extra large (30%) using a combination of UK Kennel Club breed details and personal experience. Median age was 35 months (Range 1.5–304 months).

Figure 4 shows the age distribution of all dogs. There is a marked left shift with most dogs fitting into the young/middle age category (≤7 years) (4,892) and of these 1,150 and 1,047 were ≤6 months and ≤1 year, respectively; 1,216 dogs were ≥8 years.

Figure 5 shows the Distribution of the Vet flag by age. The cut-off was 16 years (25 dogs) because in the range 17–25 years there were very few dogs (18 dogs). The Vet flag was triggered for 21% of dogs aged between 6 weeks and 6 months and thereafter there was a rise in the number of dogs for whom the flag was triggered, 31, 49, 69, and 83% for ages 4, 8, 11, and 14, respectively, with an increase in the rate of rise from 8 years.

Figure 6 gives an overview of the owners' impression of health status for the 6,108 dogs along with details of their age category and whether the Vet flag was triggered. Of the 5,002 dogs considered to be in perfect health by their owners the Vet flag was triggered for 1,343, indicating that they were not in perfect health according to their HRQL profile of scores. Of these dogs 75% were classified as young/middle aged and 25% were old.

DISCUSSION

Two vet alert models, one for young/middle aged dogs (≤7 years) and one for old dogs (≥ 8 years), were developed to identify whether dogs were sick or healthy based on four HRQL domain scores derived via an online instrument for monitoring HRQL in dogs (9). The models were created using the elastic net multivariable logistic regression with all four HRQL domain scores simultaneously included as potential variables. The elastic net method chooses which variables go into each model, with the aim of avoiding the inclusion of variables that are not predictive of the health status of the dogs. This, along with using 10-fold cross validation procedure to choose the penalty parameter, improves how well the models perform in dogs that were not part of the original data used to train the model, making the alert system more suitable for use as an early warning system outside of this initial dataset.

The variables chosen by each of the models indicated which of the HRQL domain scores were most predictive of health status.

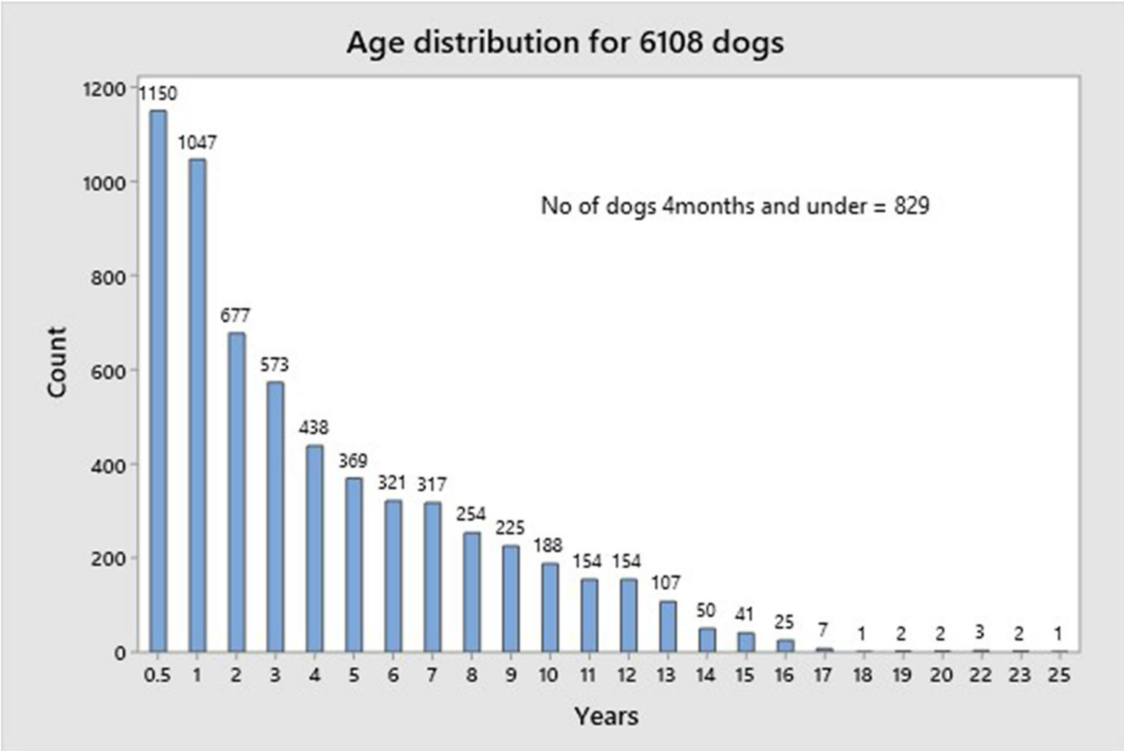


FIGURE 4 | The age distribution of all dogs.

Both the Energetic/Enthusiastic and Active/Comfortable HRQL domains were included in both models because they helped to differentiate between healthy and sick dogs in both age groups. The Happy/Content HRQL domain was only included in the model for young dogs which suggests that based on current data this HRQL domain only helps determine the health status of young dogs. It is unclear why this might be the case. It may simply be that there is not enough data to distinguish the complex relationship between the validated measures of HRQL and a veterinary diagnosis of sickness. The high correlations Happy/Content has with Energetic/Enthusiastic and Active/Comfortable within the model indicate that this may be the case, and it is these high correlations that result in Happy/Content being removed by the elastic net model's variable selection procedure. Finally, both vet alert models did not contain the Calm/Relaxed HRQL domain score. The Calm/Relaxed HRQL domain has previously been shown to be less consistent in differentiating health status compared to other domains, and this has previously been attributed to the fact that this domain may reflect stable personality traits which are less sensitive to changes in health status (9, 14). The lack of all domains in the model does not necessarily mean that each of the HRQL domains is not reflective of HRQL, but rather that a combination of them is needed to predict a veterinary diagnosis of sickness. While it may seem logical to include all the validated HRQL domains regardless, it would be statistically inappropriate to do so as this could result in worse predictive quality when the model is applied to a general dog population.

In terms of sensitivity and specificity, both vet alert models performed well, with the results for both models being close to the perfect score of 1 (**Table 1** and **Figure 1**). This meant that the models were able to identify 93 and 83% of sick dogs respectively for young/middle aged and old subjects within the data set. Similarly, they correctly predicted the health status of 83% of healthy dogs for both young/middle-aged and old dogs. While these sensitivities and specificities are only reflective of the performance of the models within the dogs selected for the study, the model selection procedure which included cross-validation for selection of the variables and an elastic net regression coefficient estimation method, should ensure that similar performance will be obtained for other cohorts of dogs. Indeed, the case shown in **Figures 2A,B** of the dog with IBD provides evidence for the soundness of the alert. When only one domain score deteriorated (Happy/Content), which in the case of old dogs is not included in the alert algorithm, and besides which could have been the result of conditions other than health, there was no alert triggered, but when there was a general deterioration in scores, coinciding with a flare-up of the clinical condition, the alarm was raised.

The majority of dogs for whom there were HRQL data recorded were in the young/middle aged category with a large proportion of these aged 1 year and under and this may

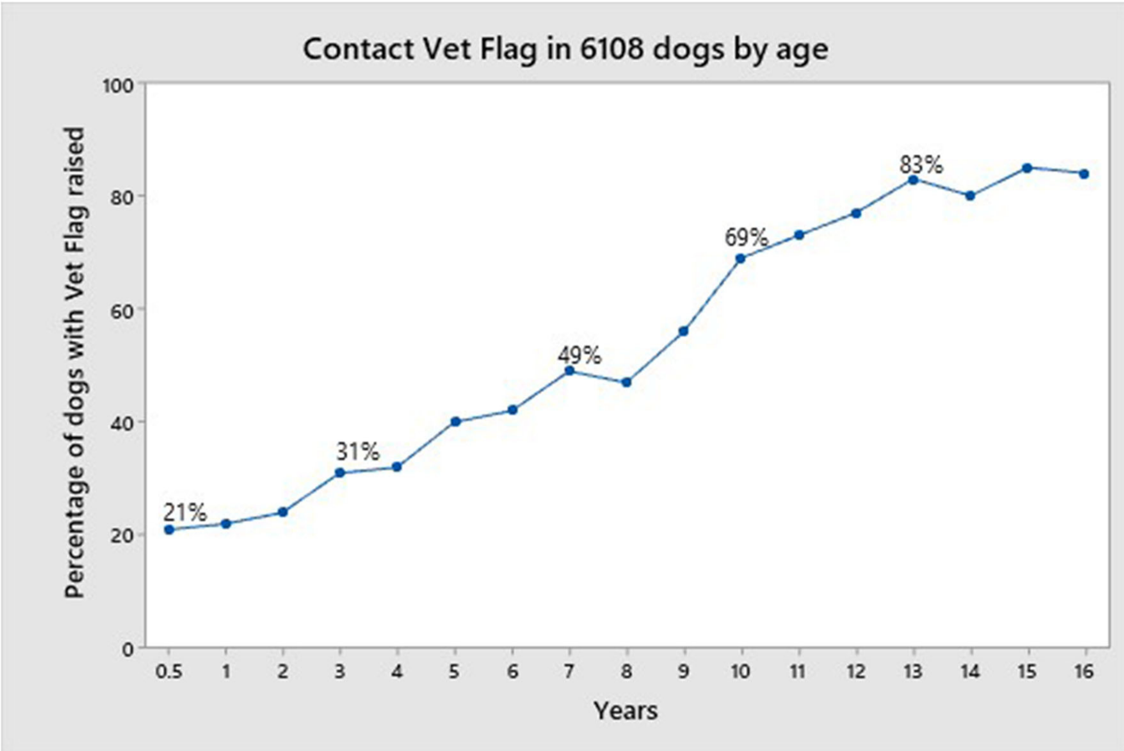


FIGURE 5 | Distribution of the Vet flag by age.

have accounted in part for the high proportion of entire dogs. Furthermore, ∼20% of this age group triggered the vet flag. While this is understandable given that some of these dogs might have had an acute illness or something as simple as a cut pad that would affect their mobility and hence their HRQL scores on the day of assessment, it is notable that 38% of these dogs were 4 months and under, and so such an explanation is less likely. The behavior patterns of young puppies differ from that of the adult and as the algorithm for the vet flag was derived using a population of dogs aged 6 months and over, this may account for the unexpected incidence of the Vet flag triggering in this group of very young dogs. Other than for the very young dogs, the incidence of vet flag triggering was as expected with a steady rise as age increased. From 1 to 7 years of age there was a 28% rise and then in a similar period from 8 to 14 years the rise was 34%. Thereafter the rate of vet flag was static at around 83%. This accelerated incidence from 8 years on concurs with the increase of chronic disease in geriatric populations, with a consequent decrease in QOL.

In people, wellness is often seen as a critical concept for understanding preventive aspects of disease, disability, and social breakdown and it has been the focus of much research relating to the support of the elderly living independently at home (18). To the authors' knowledge there is no such wellness monitoring available for people throughout their life cycle. However, the shorter lifespan of companion animals affords the opportunity to measure wellness using generic HRQL measures throughout life and has distinct advantages in terms of early disease detection, monitoring the efficacy of therapeutic interventions, and providing humane endpoints for individuals at end of life.

Currently, pet health care plan providers assure owners that by subscribing to a wellness health plan they are ensuring the best preventative health care for their pet. Other than vaccination and routine ectoparasite treatment, these generally include an annual or biannual veterinary examination with or without additional veterinary nurse consultations. However, between these visits the Vet is reliant on the owner to make an appointment if their pet seems to be unwell. It has been shown here that owners frequently miss the signs of ill health in their animals. The Vet flag was triggered in 1,343 (27%) of 5,002 dogs deemed healthy by their owners, and this compares well with statistics published for the cat (26%) (10). Of these 1,343 dogs, 75% were young/middle aged and 25% were old. Much anecdotal evidence exists to suggest that owners frequently confuse the signs of illness in the elderly animal with "just getting old," but this does not explain the high proportion of dogs that were in the young/middle aged category. Since this initial assessment may have coincided with an acute disease episode or a traumatic event, these cannot be ruled out as a cause of low scores in HRQL, but such dogs are unlikely to be deemed in perfect health by their owners. Accordingly, we may assume that the majority might have been suffering from an underlying disease process like osteoarthritis. Osteoarthritis very often begins in early life, but the clinical signs may not manifest themselves until the animal is older, or until these signs are highlighted through the use of a Health Risk Assessment administered routinely by the veterinary

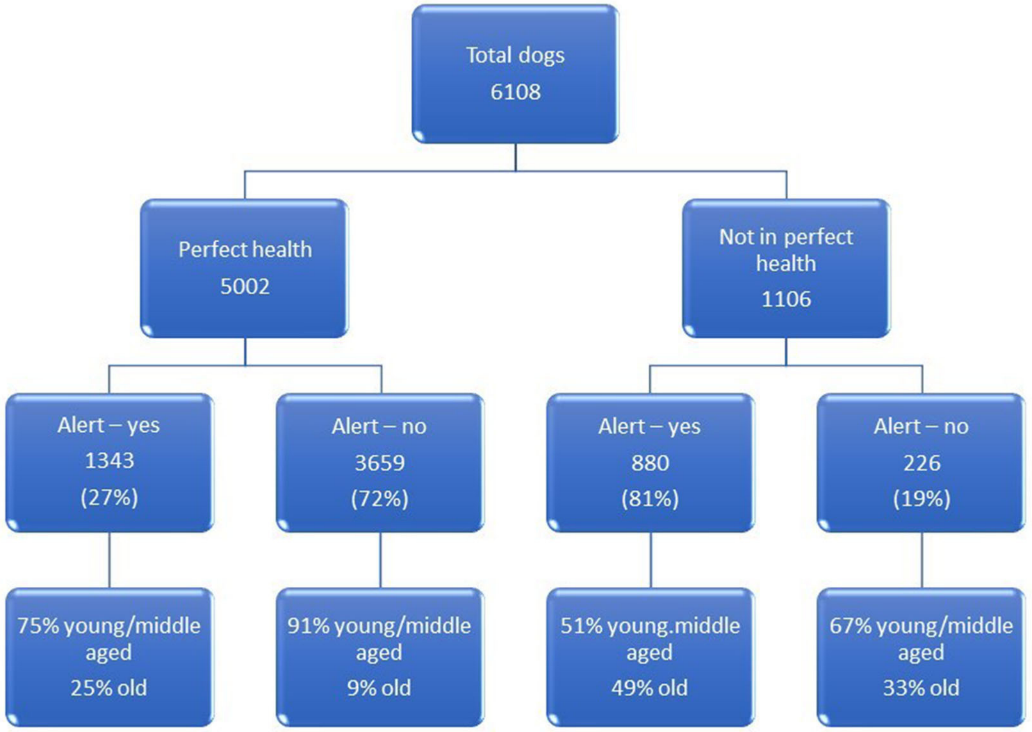


FIGURE 6 | An overview of the owners' impression of health status for the 6,108 dogs along with details of their age category and whether the Vet flag was triggered.

practice (19). If lifestyle changes/treatment are instigated early this can mitigate progression of the disease (7). Accordingly, the incorporation of a generic HRQL monitoring tool into existing wellness plans for companion animals would bridge the gap between scheduled veterinary visits. However, wellness plans involve a financial commitment for the owners which may limit their uptake to those who have the necessary resources, especially if the owner has multiple pets. Currently, to the authors' knowledge, there is no provision for wellness monitoring within any veterinary health plan. The work described here has demonstrated that the PetDialog™ app and Vet Support+ ecosystem, which supports 24/7 remote health monitoring of pets, is an efficient way for vets to provide all their owners with the opportunity to monitor their animal's wellness throughout their lifetime, providing the vet with a mechanism to identify health problems early via the vet flag and most importantly to stimulate owners to be more proactive in seeking veterinary attention. However, it was a major limitation of this study that there was no available mechanism to determine how many owners responded to the vet flag and did subsequently consult their vet. Establishing this will be important moving forward and it may be that it will be necessary to make more explicit the message that the owner receives. When the "thank you" page was designed the research team did not want the alert message for the owner to be too alarmist and so it errs on the gentle side.

It is encouraging to note that most dogs were registered for PetDialog when they were young, perhaps indicating that owners were prepared to use the app rather like a wellness plan. However, this would depend on continued use of the tool and experience with human healthcare apps has shown that to be effective continuous use of mobile Health apps is vital. Further research into the longitudinal use of VetMetrica™ within PetDialog™ is being undertaken to establish if owners are engaged in a sustainable fashion and this will be the focus of a further publication.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

Conceptualization: MW-O, JR, VD, and ES. Provision of data: JR and AW. Statistical analysis: VD and ES. Algorithm: VD, MW-O, and JR; App development: JR and MW-O. Writing, review, and editing: JR, VD, ES, MW-O, and AW. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

The authors are grateful to all the veterinary surgeons and dog owners who participated in those studies that provided data for this work and also to those owners completing assessments for their dogs on the PetDialog™ App. We are particularly grateful to Katie Wallace and Jill Thompson of the Zoetis Centre for Digital Innovation (CDI) for their invaluable assistance in supporting the collection, analysis, and validation of data through the PetDialog™ App for this study. We also acknowledge with thanks Davies Veterinary Specialists for allowing use of their data for **Figure 2**.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2020.575795/full#supplementary-material

REFERENCES

1. Shapiro M, Johnston D, Wald J, Mon D. *Patient-Generated Health Data: White Paper Prepared for the Office of the National Coordinator for Health IT by RTI International*. (2012) Available online at http://www.rti.org/pubs/patientgeneratedhealthdata.pdf
2. Volk JO, Felsted KE, Thomas JG. Executive summary of the Bayer veterinary care usage study. *J Am Vet Med Assoc.* (2011) 238:1275–82. doi: 10.2460/javma.238.10.1275
3. Banfield Pet Hospital. *State of Pet Health 2011 Report*. Portland, Ore: Banfield Pet Hospital (2011). Available online at: www.stateofpethealth.com/Content/pdf/State_of_Pet_Health_2011.pdf (Accessed May 4, 2020).
4. Banfield Pet Hospital. *State of Pet Health 2012 Report*. Portland, Ore: Banfield Pet Hospital (2012). Available online at: www.stateofpethealth.com/Content/pdf/State_of_Pet_Health_2012.pdf (Accessed May 4, 2020).
5. Banfield Pet Hospital. *State of Pet Health 2013 Report*. Portland, Ore: Banfield Pet Hospital (2013). Available online at: www.stateofpethealth.com/Content/pdf/Banfield-State-of-Pet-Health-Report_2013.pdf (Accessed May 4, 2020).
6. Fatjo J, Ruiz-de-la-Torre JL, Manteca X. The epidemiology of behavioural problems in dogs and cats: a survey of veterinary practitioners. *Anim Welf.* (2006) 15:179–85.
7. Spofford N, Lefebvre SL, McCune S, Niel L. Should the veterinary profession invest in developing methods to assess quality of life in healthy dogs and cats?. *J Am Vet Med Assoc.* (2013) 243:952–6. doi: 10.2460/javma.243.7.952
8. Reid J, Wiseman-Orr ML, Scott EM, Nolan AM. Development, validation and reliability of a web-based questionnaire to measure health-related quality of life in dogs. *J Small Anim Pract.* (2013) 54:227–33. doi: 10.1111/jsap.12059
9. Reid J, Wiseman-Orr L, Scott M. Shortening of an existing generic online health-related quality of life instrument for dogs. *J Small Anim Pract.* (2018) 59:334–42. doi: 10.1111/jsap.12772
10. Noble CE, Wiseman-Orr LM, Scott ME, Nolan AM, Reid J. Development, initial validation and reliability testing of a web-based, generic feline health-related quality-of-life instrument. *J Feline Med Surg.* (2019) 21:84–94. doi: 10.1177/1098612X18758176
11. Fayers PM, Hand DJ. Causal variables, indicator variables and measurement scales: an example from quality of life. *J R Stat Soc Ser A.* (2002) 165:233–53. doi: 10.1111/1467-985X.02020
12. Schünemann HJ, Akl EA, Guyatt GH. Interpreting the results of patient reported outcome measures in clinical trials: the clinician's perspective. *Health Qual Life Outcomes.* (2006) 4:62. doi: 10.1186/1477-7525-4-62
13. Lam CL, Lauder IJ, Lam TP, Gandek B. *Population Based Norming of the Chinese (HK) Version of the SF-36 Health Survey*. Hong Kong Practitioner (1999).
14. Davies V, Reid J, Wiseman-Orr ML, Scott EM. Optimising outputs from a validated online instrument to measure health-related quality of life (HRQL) in dogs. *PLos One.* (2019) 14:e0221869. doi: 10.1371/journal.pone.0221869
15. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc Ser B.* (2005) 67:301–20. doi: 10.1111/j.1467-9868.2005.00503.x
16. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw.* (2010) 33:1–22. doi: 10.18637/jss.v033.i01
17. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning*. New York, NY: Springer (2009).
18. Huh J, Le T, Reeder B, Thompson HJ, Demiris G. Perspectives on wellness self-monitoring tools for older adults. *Int J Med Inform.* (2013) 82:1092–3. doi: 10.1016/j.ijmedinf.2013.08.009
19. Wright A, Amodie D, Cernicchiaro N, Lascelles B, Pavlock A. PVM1 Diagnosis and treatment rates of osteoarthritis in dogs using a Health Risk Assessment (HRA) or Health questionnaire for osteoarthritis in general veterinary practice. *Value Health.* (2019) 22:S387. doi: 10.1016/j.jval.2019.04.1886

Conflict of Interest: JR is a Director in NewMetrica Research and AW is a Director in Outcomes Research International, Zoetis.

The remaining 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.

frontiers in Veterinary Science

BRIEF RESEARCH REPORT
published: 05 October 2020
doi: 10.3389/fvets.2020.576942

Assessment of Animal-Based Pig Welfare Outcomes on Farm and at the Abattoir: A Case Study

Dayane Lemos Teixeira[1,2*], *Laura C. Salazar*[2], *Daniel Enriquez-Hidalgo*[2,3,4] and *Laura A. Boyle*[5]

[1] Instituto de Ciencias Agroalimentarias, Animales y Ambientales (ICA3), Universidad de O'Higgins, San Fernando, Chile, [2] Departamento de Ciencias Animales, Pontificia Universidad Católica de Chile, Santiago, Chile, [3] Bristol Veterinary School, University of Bristol, Langford, United Kingdom, [4] Sustainable Agriculture Sciences Department, Rothamsted Research, North Wyke, Okehampton, United Kingdom, [5] Pig Development Department, Teagasc Animal and Grassland Research and Innovation Centre, Fermoy, Ireland

OPEN ACCESS

Edited by:
Alan G. McElligott,
University of Roehampton London,
United Kingdom

Reviewed by:
Mette S. Herskin,
Aarhus University, Denmark
Moira Harris,
Consultant, Shrewsbury,
United Kingdom

***Correspondence:**
Dayane Lemos Teixeira
dadaylt@hotmail.com

Specialty section:
This article was submitted to
Animal Behavior and Welfare,
a section of the journal
Frontiers in Veterinary Science

Received: 27 June 2020
Accepted: 14 August 2020
Published: 05 October 2020

Citation:
Teixeira DL, Salazar LC,
Enriquez-Hidalgo D and Boyle LA
(2020) Assessment of Animal-Based
Pig Welfare Outcomes on Farm and at
the Abattoir: A Case Study.
Front. Vet. Sci. 7:576942.
doi: 10.3389/fvets.2020.576942

The study assessed the prevalence of animal-based pig welfare outcomes on one Chilean farm and one abattoir. A total of 198 pens of slaughter pigs (9,049 pigs) were observed on farm and 54 batches (8,843 pigs) were observed at the abattoir. All assessments were conducted from outside the pen on farm and from outside the corridor where pigs were unloaded from the truck at the abattoir. Batch size and number of pigs with ear, tail and skin lesions, hernias, rectal prolapse, bursitis, and lameness were recorded. Data were analyzed using generalized linear mixed models. There was a large variation among pens on farm and among batches at the abattoir for all outcomes. Bursitis was the most prevalent outcome recorded in both locations, followed by ear lesions recorded on farm and by tail lesions recorded at the abattoir. Ear lesions' prevalence was higher on farm ($P < 0.001$), while tail lesions, hernia, and bursitis prevalence were higher at the abattoir ($P < 0.001$). Ear lesions' prevalence on farm was higher in female and mixed-sex groups than in male groups ($P < 0.01$), but male groups tended to have a higher tail lesions' prevalence ($P < 0.1$). The results show a difference in welfare outcomes, suggesting that assessment of outcomes on farm could complement ante-mortem inspections at the abattoir. However, as the same animals were not inspected in the two locations and there is the possibility of a seasonal influence on the results, the findings should be interpreted with caution and further research is required.

Keywords: pig, welfare, lesion, inspection, ante-mortem (AM)

INTRODUCTION

The primary function of ante- and post-mortem inspection is the protection of public health by ensuring food safety (1). As part of this process, the detection of illness or injuries during ante-mortem and lesions during the post-mortem inspection can lead to whole or partial condemnation of carcasses. There is considerable variation in the amount and quality of ante- and post-mortem data available internationally (2). However, it can play a valuable role in reducing financial losses (3, 4) and better informing herd health and welfare management plans (5). These data are routinely collected for disease surveillance (1), but they are also used in epidemiological studies to investigate risk factors (6), farm performance indicators (7), geographical or seasonal differences (8), and variation between herds (9). Furthermore, there is growing interest in the collection of information relating to welfare of animals at meat inspection (10, 11).

Nonetheless, death or euthanasia of severely affected pigs before slaughter could mean an underrepresentation of important data relevant to pig welfare collected at the abattoir. Hence, although meat inspection carried out at the abattoir offers an ideal opportunity for continuous and practical measurement of health and lesions, the prevalence of lesions detected is unlikely to be an exact representation of the extent of the problem on farm. Furthermore, some researchers have concerns about the probability of detecting anomalies during routine ante-mortem inspection at the abattoirs (12–14), arguing that time constraints, overcrowding, poor lighting, soiled hides, smell and noise pose challenges (14). On farm welfare assessments are commonly done daily and before loading [e.g., (15)] and could help overcome concerns related to ante-mortem inspection at the abattoirs. Importantly, on farm inspection can help with segregation of pigs that are at high-risk for gross abnormalities (13, 16). For example, suspect pigs have higher risk of suffering transport injuries than normal pigs (13); therefore, on farm pre-selection of suspect pigs could facilitate (rather than replace) abattoir-based ante-mortem inspection and reduced meat loss due to increased risk for injuries by separating pigs prior to transport into groups with and without visible lesions (17). Finally, sick and injured animals could be detected in advance, with the potential to prevent animals that are not fit for slaughter or for transport being sent to the abattoir (5). The aim of the current study was to assess animal-based pig welfare outcomes on farm and the abattoir.

MATERIALS AND METHODS

This was an observational study whereby pigs were managed according to routine practices on one Chilean commercial farm and in one commercial abattoir. The farm (May to October 2018) and the abattoir (October 2018 to May 2019) were visited over 8 and 34 days, respectively, and a welfare assessment of slaughter pigs was conducted. Animals inspected at the abattoir belonged to the same farm, but the same animals were not inspected in the two locations. Each visit (on farm and at the abattoir) lasted between 5 and 7 h.

A total of 198 pens of female and castrated male slaughter pigs were observed on farm accounting for a total of 9,049 pigs. The average group size was 45.7 (*SD* 4.85 pigs/pen; range 17–54); most of the pigs were kept on farm in single sex groups: 42.9% were female, 47% were male, and 10.1% were mixed-sex groups. Pens were selected on farm using proportionate stratified sampling to ensure that different environmental and pen characteristics were equally represented. Hospital pens on farm were excluded to ensure that pens were representative of the general population of the farm.

At the abattoir, 54 batches were observed accounting for a total of 8,843 pigs. The average batch size was 163.8 (SD 39.74 animals/batch; range 25–200). Pigs were mixed before being transported and arrived at the abattoir in mixed-sex groups. Transport system was the same for all batches. All animals observed were tail docked (5–10 cm) following normal management at Chilean farms. Batches evaluated were part of a larger experiment that aimed to investigate the association between ante- and post-mortem inspection welfare outcomes of slaughter pigs. Animals were transported and handled according to normal routine. The transports were conducted between the farm and the slaughterhouse 120 km away, accessed by motorways and secondary roads. The journey took approximately 2 h and 30 min.

TABLE 1 | Animal-based welfare outcomes recorded during welfare assessment on a commercial farm and a commercial abattoir [adapted from (18)].

Outcome	Description
Ear lesions	Bloody, swollen, and/or amputated ear
Tail lesions	Bloody, swollen, and/or amputated tail
Skin lesions	Presence of deep wound and/or hematoma
Hernia	Umbilical, scrotal, and/or inguinal hernia
Rectal prolapse	Internal tissue extrudes from the rectum
Bursitis	Presence of inflamed bursae (tangerine-sized or larger) on at least one limb(s)
Lameness	Very reluctant to walk, minimal weight-bearing on the affected limb or inability to move

Only severe cases of ear, tail and skin lesions and lameness were considered.

Two observers (a veterinarian and a veterinary technician) were trained at the beginning of the study to ensure inter-observer reliability. All assessments were conducted from outside the pen on farm and from outside the corridor where pigs were unloaded from the truck at the abattoir. This was in accordance with how stock people and veterinary inspectors commonly inspected pigs at the two locations [adapted from (18)]. The batch size and the number of pigs with ear lesions, tail lesions, skin lesions, hernias, rectal prolapse, bursitis and lameness [adapted from (18)] were recorded (**Table 1**). Only severe cases of ear, tail and skin lesions and lameness were considered. Unfortunately, it was not possible to assess both sides of the pigs' body to detect the presence of certain animal-based welfare outcomes (e.g., skin lesions, bursitis), both on farm and at the abattoir assessments.

Statistical Analysis

Animal-based welfare outcomes were recorded as the number of pigs affected per pen or batch and expressed as the percentage of pigs affected for each outcome. The median and interquartile range (IQR) of the prevalence of the different animal-based welfare outcomes were calculated for farm and abattoir assessments and ranked to identify the most prevalent outcomes within the two measurements. The number of pens where at least one animal was affected by each animal-based welfare outcome was identified to evaluate whether certain outcomes might be lowly prevalent but spread over the farm. Due to the very low prevalence of skin lesions and rectal prolapse they were not analyzed further. All statistical analyses were conducted using SAS 9.3 (SAS Institute, Inc., Cary, NC, USA). The effect of assessment location and animal sex on the prevalence of each separate animal-based welfare outcome were analyzed using generalized linear mixed models (Proc Glimmix). Assessment

TABLE 2 | Number and percentage (%) of pens on farm (198 pens; 9,049 pigs) and batches at the abattoir (54 batches; 8,843 pigs).

	On farm				At the abattoir			
	Pens		Prevalence		Batches		Prevalence	
	n	%	Median %	IQC %	*n*	%	Median %	IQC %
Ear lesions	74	37.4	0.0	0.0–2.3	11	20.4	0.0	0.0–0.0
Tail lesions	31	15.7	0.0	0.0–0.0	40	74.1	2.8	0.0–9.44
Skin lesions	0	0.0	0.0	0.0–0.0	12	22.2	0.0	0.0–0.0
Hernia	25	12.6	0.0	0.0–0.0	22	40.7	0.0	0.0–0.6
Rectal prolapse	2	1.0	0.0	0.0–0.0	6	11.1	0.0	0.0–0.0
Bursitis	170	85.9	6.5	4.1–9.3	54	100.0	11.7	8.3–13.9
Lameness	24	12.1	0.0	0.0–0.0	6	11.1	0.0	0.0–0.0

With at least one pig affected by animal-based welfare outcomes, including the median prevalence and interquartile range (IQR) of pigs affected per pen and per batch (%) of each animal-based welfare outcome observed on the two locations.

TABLE 3 | Effect of assessment location and animal sex on the prevalence of animal-based welfare outcomes in slaughter pigs assessed on farm (198 pens; 9,049 pigs) and at the abattoir (54 batches; 8,843 pigs).

	Location			Sex[a]			
	Farm	Abattoir	*P*-value	Female	Male	Mixed	*P*-value
Ear lesions	2.1 (1.92–2.31)	−0.3 (−0.77 to 0.23)	***	2.3 (2.05–2.53)[a]	1.3 (0.94–1.70)[b]	2.2 (1.73–2.71)[a]	**
Tail lesions	1.5 (1.11–1.84)	2.5 (2.14–2.78)	***	1.9 (1.38–2.46)	1.3 (0.98–1.70)	0.9 (0.24–1.60)	+
Hernia	0.9 (0.74–1.17)	0.2 (−0.03 to 0.42)	***	1.0 (0.80–1.21)	0.9 (0.62–1.13)	0.9 (0.38–1.46)	ns
Bursitis	2.0 (1.96–2.11)	2.4 (2.31–2.59)	***	2.1 (1.96–2.21)	2.0 (1.85–2.09)	2.1 (1.84–2.37)	ns
Lameness	1.1 (0.88–1.38)	1.1 (0.65–1.64)	ns	1.3 (0.94–1.59)	0.8 (0.36–1.19)	1.3 (0.62–1.98)	ns

Least squares means (95% CI) of models with gamma distribution are presented.
*[a]Abattoir assessments were excluded as sex of pigs were not assessed in this location; ns, not significant; +, Tendency (P < 0.1); **P < 0.01; ***P < 0.001.*

location ($n = 252$) and animal sex ($n = 198$; abattoir assessments excluded) were included in the models as fixed effects and gamma as the type of distribution. Least squares means (95% CI) are presented. Statistical effects and tendencies were reported when $P < 0.05$ and $P < 0.10$, respectively.

RESULTS

Prevalence of Animal-Based Welfare Outcomes Assessed on Farm and at the Abattoir

The number and percentage (%) of pens on farm ($n = 198$) and batches at the abattoir ($n = 54$) with at least one pig affected by animal-based welfare outcomes is shown in **Table 2**. In general, a large variation was observed among pens on farm and among batches at the abattoir for all animal-based welfare outcomes. Bursitis was the most prevalent outcome recorded in both locations, followed by ear lesions recorded on farm and by tail lesions recorded at the abattoir.

Effect of Assessment Location and Animal Sex

The effect of assessment location and animal sex on the prevalence of animal-based welfare outcomes assessed on farm and at the abattoir is presented in **Table 3**. Detected prevalence of ear lesions was higher on farm compared to the abattoir assessments ($P < 0.001$). In contrast, detected prevalence of tail lesions, hernia and bursitis were higher at the abattoir than on farm ($P < 0.001$). Within on farm assessments, the prevalence of ear lesions was higher in female and mixed-sex groups than in male groups ($P < 0.01$), but male groups tended to have a higher prevalence of tail lesions than the other groups ($P < 0.1$).

DISCUSSION

The current case study provides results of animal-based welfare outcomes assessed in slaughter pigs on farm and at the abattoir. In general, the detected prevalence of animal-based welfare outcomes assessed in both locations was very low, which is positive for pig welfare. However, considering that only one commercial farm supplying one abattoir participated in the study, the findings should be interpreted with caution and does not represent a general or national figure. Also, it is important to highlight that the same animals were not inspected in the two locations, which was a major limitation in the methodology.

Our findings show a wide variation in the prevalence of animal-based welfare outcome measures among pens on farm and among batches at the abattoir, which could suggest that different risk factors on farm and during transit might play a role and may contribute to the variation observed. Previous studies reported large variation in the prevalence of welfare outcomes between different ages/weight groups of pigs (18–20) and between countries (10, 21). It is important to note that all animals from our study were slaughter pigs with similar age and body weight.

In accordance with other studies carried out in Spain and Ireland (18, 22), bursitis was one of the most prevalent outcomes observed, even that these authors only recorded severe cases. The prevalence of bursitis was higher at the abattoir than on farm assessments, probably because the animal legs were more visible to the observer's eye while the animals were walking down from the truck rather than while grouped together in the pens on farm. Prevalence of ear and tail lesions was lower than reported by Petersen et al. (19) in Denmark and van Staaveren et al. (18) in Ireland in finishing pigs; however, similar trends for tail lesions were found by Temple et al. (22) in growing pigs in Spain. The prevalence of ear lesions on farm was higher than tail lesions, which is in agreement with van Staaveren et al. (18); while van de Meer et al. (23) found that ear biting was scored more frequently than tail biting behavior. The higher prevalence of tail lesions reported at the abattoir than in farm assessments seems to be due to the fact that hospital pens were not included in the on farm assessments but pigs from hospital pens could have been mixed with healthy pigs prior to transportation from farm to the abattoir.

Furthermore, different challenges were faced during assessments on farm and at the abattoir. On farm animals were sometimes huddling or lying down and at the abattoir the speed of unloading was often very fast. These must be considered when comparing the findings of welfare assessments on farm with assessments conducted at the abattoir. These observations are in accordance with other studies that expressed concerns regarding the time constraints and overcrowding during ante-mortem inspection at the abattoirs (12–14).

Moreover, almost 50% of the pens assessed on farm were all male groups, in which pigs presented lower prevalence of ear lesions and tended to present a higher prevalence of tail lesions than pigs in female and mixed-sex groups. In accordance with previous studies (3, 4, 24), males are more frequently affected by tail lesions than females and this trend becomes exaggerated with increasing tail lesion severity. Females tend to perform more tail-in-mouth behavior than male pigs (25), however, to our knowledge, the sex effect on ear lesions was not reported previously. The higher prevalence of ear lesions in groups of female pigs could suggest that they have a higher propensity to bite in general and to direct biting behavior toward the opposite sex (25), which is supported by our findings where the high prevalence of ear lesions was also found in mixed-sex pens. On the other hand, the etiology of ear lesions/necrosis is not elucidated yet and the sex effect could also be associated to immune susceptibility of female pigs to subsequent bacterial infection on the damaged tissue.

The period of the year and, consequently, the season of assessments on farm and at the abattoir differed greatly in the current study, which was also a limitation in the methodology. The prevalence of lesions and diseases is known to vary with season (8, 26). There is a higher prevalence of ear biting (27) in the winter months than during summer, which could explain our findings of higher prevalence of ear lesion on farm than at the abattoir assessment. Seasonal influence on tail lesion is also reported (27, 28), suggesting that heat- or cold-stressed pigs are more prone to perform tail biting (26).

The results from our study suggest that farm based assessments could augment the information collected at the abattoir ante-mortem inspection but further research following the same group of animals longitudinally from farm to the abattoir is required to confirm such assumption. Our findings also support that it is possible to identify animals with health and welfare outcomes on farm and to transport them from farm to abattoir in a separate group (13), allowing meat inspection procedures to be made more efficient (16) and to decrease the risk of microbial cross-contamination (5). Additionally, such animal-based welfare outcomes can be incorporated with food chain information (FCI) (29, 30), enabling evidence-based risk categorization of pigs before slaughter (31).

CONCLUSION

The results from our study show a difference in animal-based welfare outcomes, suggesting that assessment of animal-based pig welfare outcomes on farm could complement ante-mortem inspections at the abattoir. However, due to the use of a convenience sample, the same animals were not inspected in the two locations and the possibility of seasonality influence on the results, the findings should be interpreted with caution and further research following longitudinally the same group of animals from farm to abattoir is requited to confirm such assumption.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

This study was part of a research project approved by the Scientific Ethics Committee for Animals and Environmental Care of the Pontificia Universidad Católica de Chile (Protocol No. 170529006). Written informed consent was obtained from the owners for the participation of their animals in this study.

AUTHOR CONTRIBUTIONS

DT, LS, LB, and DE-H contributed to the concept of the work. DT and LS initiated and designed the study and performed the experiment. DT and DE-H performed statistical analysis. DT, DE-H, and LB interpreted data. DT wrote the manuscript.

DE-H, LS and LB, contributed to the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was funded by the National Commission for Scientific and Technological Research, CONICYT, through the Programme FONDECYT, Grant No. 11170822.

ACKNOWLEDGMENTS

DT acknowledge the financial support provided by CONICYT, project number FONDECYT 11170822. Great thanks to all staff of the farm and the abattoir who always supported the development of the project. Many thanks to Patricia Camila Carrasco Gonzalez for assisting in conducting the experiment and for her help with data collection.

REFERENCES

1. Edwards DS, Johnston AM, Mead GC. Meat inspection: an overview of present practices andfuture trends. *Vet J.* (1997) 154:135–47. doi: 10.1016/S1090-0233(97)80051-2
2. Harley S, More S, Boyle L, O'Connell N, Hanlon A. Good animal welfare makes economic sense: potential of pig abattoir meat inspection as a welfare surveillance tool. *Irish Vet J.* (2012) 65:1–12. doi: 10.1186/2046-0481-65-11
3. Harley S, More S, O'Connell N, Hanlon A, Teixeira D, Boyle L. Evaluating the prevalence of tail biting and carcase condemnations in slaughter pigs in the Republic and Northern Ireland, and the potential of abattoir meat inspection as a welfare surveillance tool. *Vet Rec.* (2012) 171:621. doi: 10.1136/vr.100986
4. Harley S, Boyle L, O'Connell N, More S, Teixeira D, Hanlon A. Docking the value of pigmeat? Prevalence and financial implications of welfare lesions in Irish slaughter pigs. *Anim Welf.* (2014) 23:275–85. doi: 10.7120/09627286.23.3.275
5. EFSA. Scientific Opinion on the public health hazards to be covered by inspection of meat (swine). *EFSA J.* (2011) 9:2351. doi: 10.2903/j.efsa.2011.2351
6. Mousing J, Lybye H, Barfod K, Meyling A, Rønsholt L, Willeberg P. Chronic pleuritis in pigs for slaughter: an epidemiological study of infectious and rearing system-related risk factors. *Prevent Vet Med.* (1990) 9:107–19. doi: 10.1016/0167-5877(90)90029-H
7. van Staaveren N, Teixeira DL, Hanlon A, Boyle LA. Pig carcass tail lesions: the influence of record keeping and the relationship with farm performance parameters. *Animal.* (2016) 1:1–7. doi: 10.1017/S1751731116001117
8. Elbers ARW, Tielen MJM, Snijders JMA, Cromwijk WAJ, Hunneman WA. Epidemiological studies on lesions in finishing pigs in the Netherlands. I prevalence, seasonality and interrelationship. *Prevent Vet Med.* (1992) 14:217–31. doi: 10.1016/0167-5877(92)90018-B
9. Maes D, Deluyker H, Verdonck M, Castryck F, Miry C, Vrijens B, et al. Herd factors associated with the seroprevalences of four major respiratory pathogens in slaughter pigs from farrow-to-finish pig herds. *Vet Res.* (2000) 31:313–27. doi: 10.1051/vetres:2000122
10. Cleveland-Nielsen A, Christensen G, Ersbøll AK. Prevalences of welfare-related lesions at post-mortem meat-inspection in Danish sows. *Prevent Vet Med.* (2004) 64:123–31. doi: 10.1016/j.prevetmed.2004.05.003
11. EFSA. Scientific Opinion on the use of animal-based measures to assess welfare in pigs. *EFSA J.* (2012) 10:2. doi: 10.2903/j.efsa.2012.2774
12. Petersen HH, Enøe C, Nielsen EO. Observer agreement on pen level prevalence of clinical signs in finishing pigs. *Prevent Vet Med.* (2004) 64:147–56. doi: 10.1016/j.prevetmed.2004.05.002
13. Jackowiak J, Kiermeier A, Kolega V, Missen G, Reiser D, Pointon A. Assessment of producer conducted antemortem inspection of market pigs in Australia. *Austr Vet J.* (2006) 84:351–7. doi: 10.1111/j.1751-0813.2006.00045.x
14. Schemann A, Hernández-Jover M, Hall W, Holyoake P, Toribio J. Assessment of current disease surveillance activities for pigs post-farmgate in New South Wales. *Austr Vet J.* (2010) 88:75–83. doi: 10.1111/j.1751-0813.2009.00543.x
15. Temple D, Llonch P, Mainau E, Manteca X. On-farm and post-mortem health assessment. In: Špinka M, editor. *Advances in Pig Welfare*. Duxford: Elsevier (2018). p. 357–79. doi: 10.1016/B978-0-08-101012-9.00016-2
16. Harbers A, Elbers A, Geelen A, Rambags P, Snijders J. Preselection of finishing pigs on the farm as an aid for meat inspection. *Vet Q.* (1992) 14:46–50. doi: 10.1080/01652176.1992.9694327
17. Harbers A, Snijders J, Smeets J, Blocks G, van Logtestijn J. Use of information from pig finishing herds for meat inspection purposes. *Vet Q.* (1992) 14:41–5. doi: 10.1080/01652176.1992.9694326
18. van Staaveren N, Díaz JAC, Manzanilla EG, Hanlon A, Boyle LA. Prevalence of welfare outcomes in the weaner and finisher stages of the production cycle on 31 Irish pig farms. *Irish Vet J.* (2018) 71:9. doi: 10.1186/s13620-018-0121-5
19. Petersen HH, Nielsen E, Hassing A, Ersbøll AK, Nielsen JP. Prevalence of clinical signs of disease in Danish finisher pigs. *Vet Rec.* (2008) 162:377–82. doi: 10.1136/vr.162.12.377
20. Temple D, Courboulay V, Manteca X, Velarde A, Dalmau A. The welfare of growing pigs in five different production systems: assessment of feeding and housing. *Animal.* (2012) 6:656–67. doi: 10.1017/S1751731111001868
21. Goossens X, Sobry L, Ödberg F, Tuyttens F, Maes D, De Smet S, et al. A population-based on-farm evaluation protocol for comparing the welfare of pigs between farms. *Anim Welf.* (2008) 17, 35–41.
22. Temple D, Dalmau A, Ruiz de la Torre JL, Manteca X, Velarde A. Application of the Welfare Quality® protocol to assess growing pigs kept under intensive conditions in Spain. *J Vet Behav Clin Appl Res.* (2011) 6:138–49. doi: 10.1016/j.jveb.2010.10.003
23. van der Meer Y, Gerrits WJ, Jansman AJ, Kemp B, Bolhuis JE. xA link between damaging behaviour in pigs, sanitary conditions, and dietary protein and amino acid supply. *PloS ONE.* (2011) 12:e0174688. doi: 10.1371/journal.pone.0174688
24. Valros A, Ahlström S, Rintala H, Häkkinen T, Saloniemi H. The prevalence of tail damage in slaughter pigs in Finland and associations to carcass condemnations. *Acta Agric Scand Sec A Anim Sci.* (2004) 54:213–9. doi: 10.1080/09064700510009234
25. Schrøder-Petersen DL, Simonsen H, Lawson L. Tail-in-mouth behaviour among weaner pigs in relation to age, gender and group composition regarding gender. *Acta Agric Scand Sec A Anim Sci.* (2003) 53:29–34. doi: 10.1080/09064700310002017
26. Schrøder-Petersen DL, Simonsen HB. Tail-biting in pigs. *Vet J.* (2001) 162:196–210. doi: 10.1053/tvjl.2001.0605
27. Blackshaw JK. Some behavioural deviations in weaned domestic pigs: persistent inguinal nose thrusting, and tail and ear biting. *Anim Prod.* (1981) 33:325–32. doi: 10.1017/S000335610003172X
28. Penny RH, Hill F. Observations of some conditions in pigs at the abattoir with particular reference to tail biting. *Vet Rec.* (1974) 94:174. doi: 10.1136/vr.94.9.174
29. Felin E, Jukola E, Raulo S, Heinonen J, Fredriksson-Ahomaa M. Current food chain information provides insufficient information

for modern meat inspection of pigs. *Prevent Vet Med.* (2016) 127:113–20. doi: 10.1016/j.prevetmed.2016.03.007

30. Felin E, Hälli O, Heinonen M, Jukola E, Fredriksson-Ahomaa M. Assessment of the feasibility of serological monitoring and on-farm information about health status for the future meat inspection of fattening pigs. *Prevent Vet Med.* (2019) 162:76–82. doi: 10.1016/j.prevetmed.2018.11.009
31. Buncic S, Alban L, Blagojevic B. From traditional meat inspection to development of meat safety assurance programs in pig abattoirs–The European situation. *Food Control.* (2019) 106:106705. doi: 10.1016/j.foodcont.2019.06.031

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.

frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 13 November 2020
doi: 10.3389/fvets.2020.564201

Therapy Dogs' and Handlers' Behavior and Salivary Cortisol During Initial Visits in a Complex Medical Institution: A Pilot Study

Stephanie D. Clark, Jessica M. Smidt and Brent A. Bauer*

Division of General Internal Medicine, Section of Integrative Medicine and Health, Mayo Clinic, Rochester, MN, United States

OPEN ACCESS

Edited by:
Alan G. McElligott,
University of Roehampton London,
United Kingdom

Reviewed by:
Federica Pirrone,
University of Milan, Italy
Edward Narayan,
The University of
Queensland, Australia
Cassandra Raby,
University of Liverpool,
United Kingdom

***Correspondence:**
Stephanie D. Clark
sdclark4@icloud.com

Specialty section:
This article was submitted to
Animal Behavior and Welfare,
a section of the journal
Frontiers in Veterinary Science

Received: *20 May 2020*
Accepted: *29 September 2020*
Published: *13 November 2020*

Citation:
Clark SD, Smidt JM and Bauer BA (2020) Therapy Dogs' and Handlers' Behavior and Salivary Cortisol During Initial Visits in a Complex Medical Institution: A Pilot Study. Front. Vet. Sci. 7:564201. doi: 10.3389/fvets.2020.564201

Therapy dogs provide health benefits for individuals who suffer from illnesses, such as dementia, depression, loneliness, and aggression. Therapy dogs' impact on human health has been thoroughly studied; however, studies on dog welfare have been limited. Additionally, as dogs have evolved with humans, they have learned to read non-verbal social cues. Dogs can read humans' non-verbal body language and can react to their emotions. However, the body language of dogs is poorly understood and can lead to dog owner-directed aggression. Communication plays a vital role to be a cohesive therapy team. The purpose of this study was to assess perceived stress and cortisol concentrations in therapy dogs and their handlers during the first three visits in a hospital setting. Moreover, the study aimed to investigate whether, while in an overstimulating environment, a therapy dog handler can observe his or her dog's body language and correlate such observations to the dog's stress. Nine therapy dog teams from Mayo Clinic's Caring Canine Program participated in this study. A baseline salivary cortisol was collected from the handler and therapy dog each day of the visits. Once the team arrived, a pre-visit salivary cortisol was collected from the handler and therapy dog and, afterward, a post-visit salivary cortisol. Handlers were also asked to fill out a perceived stress survey on their own stress and that of their therapy dogs'. Behavior was documented by a staff member and the handler. For each visit, the therapy dogs were at the hospital on average 47 min and visited with nine people. There was significant correlation ($P = 0.02$) between the owner's perceived stress of his or her therapy dog and the dog's salivary cortisol concentrations. The handlers noted medium to high stress, and those dogs had higher cortisol concentrations post-visit. There was no significant difference in salivary cortisol for the handler and therapy dog over the course of the three visits and comparing pre- and post-visit. Overall, the dogs displayed mixed behaviors, with the three most reported being panting, lip licking, and yawning. However, salivary cortisol results suggest that the handlers and therapy dogs maintained their welfare state throughout the visits.

Keywords: animal welfare, behavior, communication, cortisol, therapy dogs

INTRODUCTION

Dogs have been an integral part of humans' lives since the early Paleolithic time; however, the exact origin and date remain relatively vague (1). Not only have dogs created companions, but they have been noted to assist psychologically and emotionally. According to the Alliance of Therapy Dogs (2), as far back to the ancient Greeks, animals have been used to assist with mental and physical health. Since then, animals have been used for people who suffer from dementia, depression, loneliness, and aggression (2). It was not until the 1960s when the first research involving animal therapy was conducted by Boris Levinson (2). Since then, numerous research projects have been conducted investigating the positive benefits of therapy dogs for humans, but evaluating the animal's welfare during these sessions is limited.

Dogs have a variety of jobs to assist humans in their daily life and their health, such as service dogs, emotional support dogs, and therapy dogs. As humans and dogs have lived together, their bond has strengthened. Dogs have become better at observing human communication, verbal and non-verbal. Dog studies have observed that dogs can use non-verbal social cues from humans to achieve tasks, such as finding food (3, 4). Moreover, Kaminski et al. (5) studied facial expressions in humans. This study noted that human facial expressions are active communication attempts. Interestingly, it was observed that dogs can pick up on these small facial expressions and understand humans and react to the emotions (5).

On the other hand, it has been noted that dogs communicate mostly with their bodies and are trying to communicate with people multiple times a day (5). According to the American Society for the Prevention of Cruelty to Animals (ASPCA), there are key focal points other than overall posture and movement, such as ears, mouth, tail, hair, sweat, and ears that should be taken into consideration when reading a dog's body language (6). However, whether humans can read a dog's body language has been noted to be difficult especially for emotions, such as fear and anxiety (7). This leads to misinterpretation of what is being communicated and can lead to an escalation in emotions and actions. For example, fear and anxiety can escalate to aggression if the dog's body language is not understood (7).

Being able to read a family dog's body language is vital to avoid aggression especially toward children but is even more important for therapy dogs during therapy visits in a hospital setting. When therapy dogs are visiting in the hospital, there are various stimuli that can be stressful to the dogs. One study (8) stated that the therapy dog's handler's personality can also be a variation on how well the therapy team communicates. This study concluded that the handler's personality influences the team's performance; and agreeableness had the strongest correlation to the cooperation of the team, ability to avoid conflict, and reduced dog aggression toward the owner. Therefore, it is imperative that owners be able to accurately observe their dog's body language, listen to what their dogs are saying, and mitigate escalation of negative emotions.

Moreover, in addition to body language, cortisol has also been widely used in past literature (9–13) to assess animal welfare, especially as a non-invasive biomarker for stress in therapy dogs (14, 15). Cortisol is a hormone that aids in immune regulation and metabolism and involved in the body's stress response (16). Therefore, cortisol has been a preferred welfare biomarker to collect, especially in animal studies where behavior is observed (9–13, 17–23).

The purpose of this study was to assess perceived stress and cortisol concentrations in therapy dogs and their handlers during the first three visits in a hospital setting. Moreover, the study aimed to investigate if therapy dog handlers can observe their dog's body language while in an overstimulating environment and correlate their observations to their and their dog's salivary cortisol concentration.

MATERIALS AND METHODS

Therapy Dog Teams

A therapy dog team consists of a dog and handler who have passed a therapy dog test and are considered a registered team. All teams that applied to the Caring Canines program at Mayo Clinic Rochester that had not volunteered with another dog or in another setting were asked to participate in the study. Handlers who had visited with other dogs were excluded, and dogs who had visited anywhere else were excluded as well. The team had to be a new team with no experience at Mayo Clinic, Rochester, MN, where all visits took place.

Therapy Dogs

Nine therapy dog teams from Mayo Clinic, Rochester's Caring Canine Program, participated in this study. All dogs had passed a therapy dog examination and were healthy, up-to-date on vaccines, and not on a raw diet. The average age for the dogs was 3.7 SD ± 2.2 years. The demographics for the nine dogs were: Dog 1 was a 2-year-old male Soft Coated Wheaten Terrier registered with Pet Partners. Dog 2 was a 2.5-year-old male mixed breed registered with Alliance of Therapy Dogs. Dog 3 was a 4-year-old female Standard Poodle registered with Pet Partners. Dog 4 was a 3-year-old female mixed breed registered with Alliance of Therapy Dogs. Dog 5 was a 2-year-old male Labrador Retriever/Golden Retriever mix registered with Therapy Dogs International. Dog 6 was a 7-year-old female Australian Cattle Dog registered with Therapy Dogs International. Dog 7 was a 6-year-old male Golden Retriever registered with Alliance of Therapy Dogs. Dog 8 was a 7-year-old male Tibetan Terrier registered with Pet Partners. Dog 9 was 1.5-year-old male Standard Poodle registered with Therapy Dogs International. All dogs had a female handler except dog 9, which had a male handler. Mayo Clinic Rochester's IACUC committee approved this study (protocol A00003248-17).

Before the start of the study, the handlers were asked to fill out general information about their dog and training techniques. During this time, the handlers were provided with a detailed description of each behavior that was written by a Karen Pryor Academy certified trainer. Each therapy team, when onboarded, had set three visits with the volunteer coordinator to adjust to the therapy dog visits. Visits consisted of inpatient and outpatient scenarios and were kept under an hour. Therapy dog visit duration, area of the hospital, and patient contact

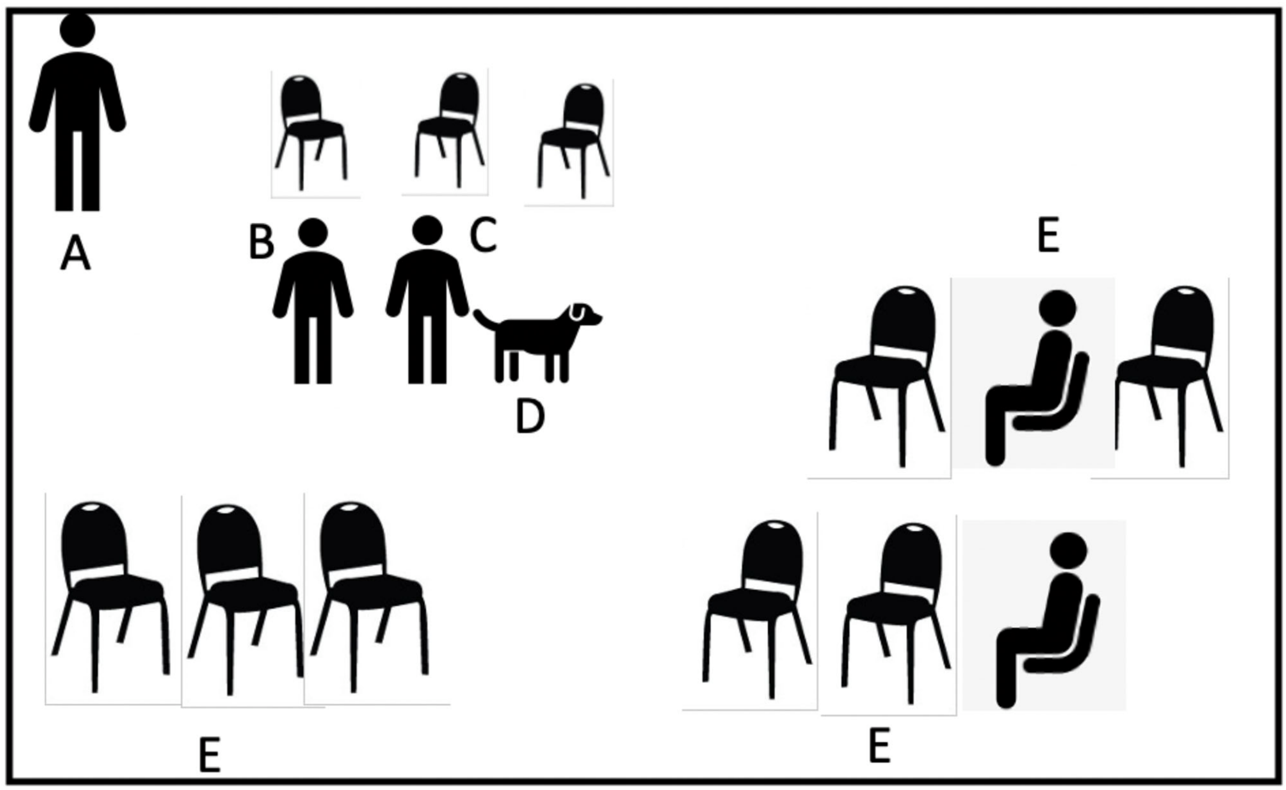


FIGURE 1 | Visit 2, outpatient waiting room. A, study staff member; B, Caring Canine's volunteer coordinator; C, therapy dog handler; D, therapy dog; E, outpatient hospital waiting room, patients have the option to approach the therapy dog if patient wanted to.

were recorded. The protocol followed the Mayo Clinic's Caring Canines protocol, including the first visit, walking around the hospital and getting accustomed to the layout and stimuli; the second visit was conducted in an outpatient waiting room; and the third visit was conducted in hospital rooms with patients. The first visit is used to get to know the campus, dogs could be touched and greeted but only in the hallways, and interactions were under 30 s. The second visit in the waiting room consisted of identifying a place to sit in the waiting room, and those interested in touching the dog could approach (**Figure 1**). The third visit was inpatient visits. The dog and handler would enter a hospital room where the patient would be either in his or her bed or sitting in the hospital chair (**Figure 2**). The dog could approach the patient and interact organically as this is how animal-assisted activities are conducted at Mayo Clinic. The team was accompanied by a study staff member to assist with saliva collection, surveys, and recorded behaviors they observed; this individual was only allowed to interact for collecting measurements. Additionally, a volunteer program leader whose role is to conduct the three sessions was present. This individual walks with the handler to direct the sessions. Only one therapy dog team was present for the visits. All therapy dog teams abided by the therapy dog policies of remaining on a 6-foot leash. Since not all therapy dog organizations allow giving treats during visits, handlers were refrained from rewarding dogs with treats for these visits. Dogs were allowed fresh water under the handler's direction. Mayo Clinic Rochester's Institutional Review Board committee approved this portion of the study (protocol 17-009412 00).

Throughout the duration of the study, the therapy dogs were not allowed to make any additional hospital visits or visit outside facilities. The day of each visit, the handler and therapy dog had saliva collected in the morning after the cortisol awakening response, before the visit, and directly after the meeting. Handlers were also asked to fill out a post-visit survey on how stressed they felt and the perceived stress of their dog. Handlers were also asked to list any behaviors noticed during the visit.

Salivary Cortisol Analysis

Salivary cortisol was used as a non-invasive procedure to assess the stress response during the visit. The morning of each visit, 30 min after the therapy dog had woken up (24), a baseline saliva sample was collected using SalivaBio's Children's Swabs, Salimetrics LLC, Carlsbad, CA (25), by the dog's handler between 0600 and 0800 h. To remain consistent, while limited to the handler's availability, all handlers were asked to schedule each session between 1,100 and 1,500 h. Another saliva sample was collected before the start of the volunteering session (pre-visit), once the dog arrived at the campus. After the visit, on average 5 min after their last interaction, a post-visit sample was collected

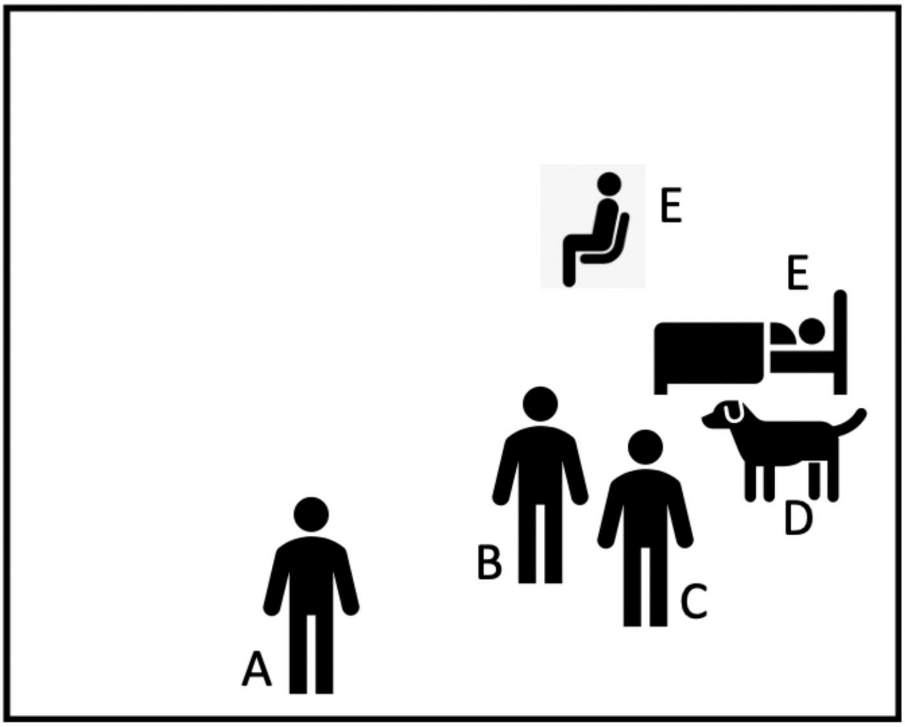


FIGURE 2 | Visit 3, inpatient hospital rooms. A, study staff member; B, Caring Canine's volunteer coordinator; C, therapy dog handler; D, therapy dog; E, inpatient hospital (could be in hospital bed or in a hospital chair).

using the same technique for saliva collection at baseline. The delay in post-visit sampling was due to walking back to an exam room for collection. Each salivary collection was under 2 min to ensure the sampling did not affect the cortisol concentration (24). A swab method was used for the dogs as passive drool was not possible for these collections. The saliva swab was placed in a swab storage tube (SST) 17 × 100 mm (25). Handler saliva samples were also collected at the same frequency of the therapy dogs; however, this was collected via passive drool. All saliva samples in the storage tubes were stored in an 80°F until shipped on dry ice overnight to Salimetrics LLC, Carlsbad, CA, to be analyzed for salivary cortisol concentrations. Salimetrics cortisol assay has a sensitivity of <0.007 μg/dl. Additionally, individual percent coefficients of variance (CV) were calculated between the duplicates for each human and dog at each time point, and the averages of the percent CV were used to calculate the intra variance CV. Human intra-assay percent CV was 4.5% and dog was 5.4%. No inter-assay CVs were calculated due to samples not being measured over different plates or days.

Surveys

A general introduction survey was administered prior to the start of the first visit. This survey contained questions regarding information on the therapy dog (i.e., breed, sex, and age), training techniques, and any training courses that may have been taken before testing to become a therapy dog. After every visit, the handler was given a modified perceived stress scale (26), which measures self-reported perceived stress and comfort level during the visits, with 0 being never/not stressed at all and 4 being very often/extremely stressed. At the end of this questionnaire, handlers were asked to rate their dog's perceived level of stress (low, medium, or high) and check off behaviors observed. Observed behaviors were combined with the study staff's observed behaviors to better capture the therapy dog's body language.

Statistical Methods

For this pilot study, the average cortisol concentrations for visit (first, second, and third) and time of day (baseline, pre-session, and post-session) for dog and human samples were reported using least square means and standard errors. The perceived survey was a Likert scale questionnaire. Likert scale questions were described using frequency percentages and analyzed pairwise for each day combination using the Wilcoxon signed rank test for repeated measures.

Separate models were used to determine if the visit or time of day had any impact on the cortisol levels for the dogs or their handlers. Each of the nine pairs of participants had their cortisol checked three times a day on each visit, resulting in a sample size of 81 cortisol results. To account for fixed and random effects in the repeated measure design, linear mixed models with an outcome of cortisol concentrations were formulated for both dogs and handlers. The time of day that the cortisol sample was tested and the visit number were included as independent

TABLE 1 | Linear mixed models for comparison of cortisol concentrations between handlers and dogs over time.

	Estimate (SE)	*F*-value	*P*-value*
Dog model			
Intercept	0.17		
Time of day		0.20	0.824
Baseline vs. Post	0.07 (0.16)		
Pre vs. Post	−0.02 (0.15)		
Visit		0.26	0.773
Day 1 vs. Day 3	0.12 (0.24)		
Day 2 vs. Day 3	0.15 (0.21)		
Human model			
Intercept	0.19		
Time of day		9.96	0.002
Baseline vs. Post	0.17 (0.04)		
Pre vs. Post	0.02 (0.04)		
Visit		0.38	0.687
Day 1 vs. Day 3	−0.05 (0.06)		
Day 2 vs. Day 3	−0.03 (0.05)		

*ar(1) covariance structure.

TABLE 2 | GEE model with low or medium/high perceived stress as the outcome.

	Odds Ratio (95% CI)*	*P*-value
Visit 1	1.12 (0.14, 9.11)	0.914
Visit 2	0.44 (0.05, 3.83)	0.460
Visit 3	ref	ref
Cortisol concentration (unit = 0.5)†	0.18 (0.04, 0.79)	0.024

*Model outcome is the odds of handler perceiving low stress rather than medium or high stress.
†Cortisol concentration at the post point for dog.

variables in the model. The ID of the dog/handler was included as a random effect in the model to control for the correlation between repeated measures. Akaike and Bayesian information criteria were used to compare covariance structures. For the linear mixed models for both handler and dog, a first-order autoregressive covariance structure was utilized.

To analyze the observed stress of the dogs from their handlers to the cortisol concentrations, generalized estimating equation (GEE) models were used to evaluate differences between visits using an outcome of low perceived stress vs. medium or high perceived stress. There was only one instance of high perceived stress, so the medium and high perceived stress answers were combined into one group. Perceived stress for each dog was collected on all 3 days for a total sample size of 27. The handler ID was included in the model as an identity variable. The independent variables of the cortisol level of the dog at the post-session time point and the visit day are included in the model. A $P < 0.05$ was considered significant in all cases. SAS (SAS version 9.4, SAS Institute Inc.) was used to analyze the data.

RESULTS

Salivary Cortisol

The therapy dogs did not have any statistical significance when comparing salivary cortisol across time of day and between visits (**Table 1**). Baseline mean salivary cortisol concentrations were 0.59 SE ± 0.24, 0.32 SE ± 0.24, and 0.10 SE ± 0.23 for visits 1, 2, and 3, respectively. Pre-visit mean salivary cortisol concentrations were 0.07 SE ± 0.24, 0.17 SE ± 0.24, and 0.54 SE ± 0.26 for visits 1, 2, and 3, respectively. Lastly, post-visit mean salivary cortisol concentrations were 0.18 SE ± 0.23, 0.32 SE ± 0.23, and 0.20 SE ± 0.24 for visits 1, 2, and 3, respectively.

There were statistical differences ($P = 0.002$) in salivary cortisol concentrations across time of day for the handlers, with baseline being the highest concentration of salivary cortisol and post-session being the lowest (**Table 1**). Baseline mean salivary cortisol concentrations were 0.30 SE ± 0.06, 0.33 SE ± 0.06, and 0.38 SE ± 0.06 for visits 1, 2, and 3, respectively. Pre-session mean salivary cortisol concentrations were 0.17 SE ± 0.06, 0.17 SE ± 0.06, and 0.23 SE ± 0.06 for visits 1, 2, and 3, respectively. Lastly, post-session mean salivary cortisol concentrations were 0.16 SE ± 0.06, 0.18 SE ± 0.06, and 0.16 SE ± 0.06 for visits 1, 2, and 3, respectively.

Survey Responses

There was significant correlation ($P = 0.02$) between the owner's perceived stress of their therapy dog and the dog's salivary cortisol concentrations (**Table 2**). The dogs' salivary cortisol concentration increased by 0.5 nmol/ml as the owner reported his or her dog experiencing medium or high stress compared to those who reported his or her dog to be low stressed during the visits.

There were no statistical differences in the perceived stress questions. However, when asked "During the visit, how often were you upset because something happened unexpectedly?" (**Supplementary Table 1**), nearly two-thirds of the handlers responded that they had never felt this way. The handlers (55.6%) responded with sometimes when asked how often they felt nervous or stressed, but the amount decreased throughout visits 2 and 3, with 11.1% responding with sometimes during each visit.

When asked "How often did you feel unable to control your dog?," the handlers felt less confident as the visits progressed, with 66.7% responding with never during the first visit, 55.6% during the second, and 44.4% during the third visit. One handler responded that he or she had felt he or she did not have control over the dog nearly the entire time, but this was only for the first visit. Moreover, as the visits progressed, 55.6% of the owners responded almost never when asked how often they had to stop the visit because their dog was not comfortable. This was from visit 1 (22.2%) and visit 2 (11.1%). Lastly, the handlers reported that they had noticed their dog showing signs of stress sometimes during visit 1 (44.4%), 2 (55.6%), and 3 (33.3%).

Observed Behavior

Over the course of the three visits, the therapy dogs were on the hospital campus for an average of 47 min (range: 25–60 min) and visited with nine people (range: 4–24 people)

(**Supplementary Table 2**). Overall, the dogs displayed a mix of behaviors with the most common (22 out of the 27 visits) being panting. The second and third most frequent behaviors observed were lip licks (19) and yawning (14). Moreover, leaning into people the therapy dog was interacting with was noted in 13 out of 27 visits. Head turning away from stimulus (10) and "wet dog" shake (9) were additional behaviors observed during the visits.

DISCUSSION

Salivary Cortisol

Salivary cortisol is used as a non-invasive measurement for assessing stress in therapy dogs (9–13). Baseline was set in correspondence with the cortisol awakening response, where the cortisol concentrations are at its peak (10). The cortisol awakening response is known to be one of the highest, naturally occurring cortisol peaks during the day due to the circadian rhythm, which provides a kick start to the production of cortisol within the first hour of waking up (10). The authors decided to use this as the baseline to determine if pre- or post-visit concentrations would exceed their natural peak. This was observed in this study's results. The baseline for all three visits had the largest cortisol concentration for the handlers and their dogs. While the pre- and post-visit cortisol concentrations were not significantly different, the post-visit concentrations were typically lower. This suggests that the handlers and dogs were not negatively affected by the visits (9–12). While previous studies have noted that handler's cortisol concentrations increased with the duration of the visit and the therapy dog's cortisol increased with the number of times they visited (10), this was not observed in the current study.

One limitation to this study is that salivary cortisol cannot be used individually for a definitive assessment of overall welfare, especially stress. However, this does give researchers a basis to further expand studies with additional parameters, such as heart rate variability, oxytocin concentrations, tympanic membrane temperature, and behavior of the dog (27–31). Moreover, handler compliance was assumed for participation during the study. The study staff explained and demonstrated the salivary collection procedure and provided details as to when to collect it; however, a study member was not present in the handler's home during the baseline collection. A second limitation was these were real sessions with new therapy dog teams and times could not be controlled for when the visit took place. To minimize the effect of the circadian rhythm on the cortisol concentrations, the authors asked that all visits be conducted between 1,100 and 1,500 h.

Survey Responses

Utilization of a perceived stress survey for handlers to evaluate emotional welfare during therapy dog visits in a hospital is a relatively new tool. Interestingly, the handlers reported for most of the questions they either felt comfortable during the visits or did not feel upset when something unexpected happened. This suggests that the required pre-therapy visit classes and preparation that must be completed prior to the three visits may be preparing the handlers appropriately. Furthermore, Haubenhofer and Kirchengast (10) reported that therapy dog handlers described emotions experienced during therapy dog sessions like those emotions experienced in their everyday life. The positive descriptors reported were "interesting," "joyful," and "power dispensing," while the negative descriptors were "stressful," "physically and emotionally encumbering," and "straining" (10).

The handlers in this study reported feeling less in control of their therapy dog as the visits progressed. While this did not align with the salivary cortisol concentrations, it is something that should be studied further to determine if this correlates to the stress increasing with the duration of the visits, which was observed by Haubenhofer and Kirchengast (10). Moreover, therapy dogs are given ratings during their therapy test, of "predictable" or "complex," which can be considered on future studies if this has any impact on how dogs handle stress. Lastly, the handlers reported noticing their therapy dog becoming stressed at some point during the visits more than a third of the time for all visits, which seems to be common during therapy dog visits, regardless if it has an overall negative effect on the therapy dog's welfare (9, 12). Further studies should evaluate the welfare of therapy dogs over time to determine if there is a long-term effect from sometimes stressful therapy visits.

Observed Behavior

While behavior can be subjective at times, it has been used as a non-invasive tool for assessing stress and overall welfare of dogs (12, 32). Of the observed behavior, the therapy dogs displayed a multitude of behaviors, with the most observed being panting, lip licks, yawning, leaning into people, turning away from a stimulus, and "wet dog" shake. Similar behaviors were observed by Glenk et al. (12). In the 2014 study, the reported behaviors were lip licks, yawning, paw lifting, body shake, tail wagging, and panting. In previous studies, lip licks and yawning have been associated with social conflict situations in dogs (33), and panting and tail wagging were associated with chronic stress (34). Additionally, lip licks and wet dog shake were correlated with higher cortisol concentrations during therapy dog visits (12). However, research has noted that these behaviors may be less related to the overall stress of the dog and more closely related to the dog coping with the stress (17, 35). A limitation of the observed behavior is relying on handlers to properly identify behaviors. The authors tried to account for this by providing handlers with a detailed description of the behaviors. Moreover, due to behavior being subjective and mixed studies on if it is related to the stress being experienced or a way of managing stressful situations, future studies should focus on quantifying behaviors and connecting them to specific events during a therapy visit. Additionally, future studies should apply the use of objective coding software (12) to analyzing behavior to provide a more reliable measurement. This was not utilized in this study due to study funding.

Conclusion

This pilot study's results suggest that the handlers and therapy dogs maintained their welfare state throughout the visits and throughout the process. The non-invasive parameters utilized in this study suggested that the handlers and therapy dogs may have even been in a better welfare state post-visit. Furthermore, the significance in the handlers' assessment of his or her dog's perceived stress in correlation with the increase in therapy dog salivary cortisol suggests that the handler in a therapy team is perceptive to the therapy dog's welfare. One of the major limitations of this study was that salivary cortisol was the only biomarker collected. Past literature has demonstrated that cortisol can increase due to exercise and general arousal (positive or negative) (13, 15, 23). It is recommended that future studies use cortisol and additional biomarkers, such as tympanic membrane temperature (36), heart rate variability (37), and salivary oxytocin (27, 38). Furthermore, future studies would be beneficial in observing different therapy dog programs and their therapy teams' perceived stress along with the addition of objective physiological parameters.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Mayo Clinic Institutional Review Board. The patients/participants provided their written informed consent to participate in this study. The animal study was reviewed and approved by Mayo Clinic Rochester's IACUC committee approved this study (protocol A00003248-17). Written informed consent was obtained from the owners for the participation of their animals in this study.

AUTHOR CONTRIBUTIONS

SC contributed to the experimental design and formal analysis. SC and JS contributed to the performance of the experiment and data curation. SC, JS, and BB contributed to writing the original draft. All authors edited and approved the final manuscript.

FUNDING

This work was supported by the Mayo Clinic CTSA through grant number UL1TR002377 from the National Center for Advancing Translational Sciences (NCATS), a component of the National Institutes of Health (NIH).

ACKNOWLEDGMENTS

This publication was made possible by John P. and Carole E. Gregory Foundation and Levitetz Family Foundation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2020.564201/full#supplementary-material

REFERENCES

1. Frantz LAF, Mullin VE, Pionnier-Capitan M, Lebrasseur O, Ollivier M, Perri A, et al. Genomic and archaeological evidence suggest a dual origin of domestic dogs. *Science*. (2016) 352:1228–31. doi: 10.1126/science.aaf3161
2. Alliance of Therapy Dogs (2018). Available online at: https://www.therapydogs.com/animal-therapy/
3. Hare B, Brown M, Williamson C, Tomasello M. The domestication of social cognition in dogs. *Science*. (2002) 298:1634–6. doi: 10.1126/science.1072702
4. Kaminski J, Nitzschner M. Do dogs get the point? a review of dog-human communication ability. *Learn Motivat*. (2013) 44:294–302. doi: 10.1016/j.lmot.2013.05.001
5. Kaminski J, Hynds J, Morris P, Waller BM. Human attention affects facial expressions in domestic dogs. *Sci Rep*. (2017) 7:12914. doi: 10.1038/s41598-017-12781-x
6. ASPCA. *7 Tips on Canine Body Language*. (2020). Available online at: https://www.aspcapro.org/resource/7-tips-canine-body-language
7. Demirbas SY, Ozturk H, Emre B, Kockaya M, Ozvardar T, Scott A. Adults' ability to interpret canine body language during a dog–child interaction. *Anthrozoös*. (2016) 29:581–96. doi: 10.1080/08927936.2016.1228750
8. Jamieson LTJ, Baxter GS, Murray PJ. Who's a good handler? Important skills and personality profiles of wildlife detection dog handlers. *Animal*. (2018) 8:222. doi: 10.3390/ani8120222
9. Clark SD, Smidt JM, Bauer BA. Welfare considerations: Salivary cortisol concentrations on frequency of therapy dog visits in an outpatient hospital setting: a pilot study. *J Vet Behav*. (2019) 30:88–91. doi: 10.1016/j.jveb.2018.12.002
10. Haubenhofer DK, Kirchengast S. Dog handlers' and dogs' emotional and cortisol secretion responses associated with animal-assisted therapy sessions. *Soc Anim*. (2006) 15:127–50. doi: 10.1163/156853007X187090
11. Glenk LM, Kothgassner OD, Stetina BU, Palme R, Kepplinger B, Baran H. Therapy dogs' salivary cortisol levels vary during animal-assisted interventions. *Anim Welfare*. (2013) 22:369–78. doi: 10.7120/09627286.22.3.369
12. Glenk LM, Kothgassner OD, Stetina BU, Palme R, Kepplinger B, Baran H. Salivary cortisol and behavior in therapy dogs during animal-assisted interventions: a pilot study. *J Vet Behav*. (2014) 9:98–106. doi: 10.1016/j.jveb.2014.02.005
13. McGowan RTS, Bolte C, Barnett HR, Perez-Camargo G, Martin F. Can you spare 15 minutes? The measurable positive impact of a 15-min petting session on shelter dog well-being. *Appl Anim Behav Sci*. (2008) 203:42–54. doi: 10.1016/j.applanim.2018.02.011
14. Cobb ML, Iskandarani K, Chinchilli VM, Dreschel NA. A systematic review and meta-analysis of salivary cortisol measurement in domestic canines. *Dom Anim Endocrinol*. (2016) 57:31–42. doi: 10.1016/j.domaniend.2016.04.003
15. Mesacova L, Kottferova L, Skurkova L, Leskova L, Kmecova L. Analysis of cortisol in dog hair a potential biomarker of chronic stress: a review. *Vet Med*. (2017) 62:363–76. doi: 10.17221/19/2017-VETMED
16. Serpell JA, Coppinger R, Fine AH, Peralta JM. Welfare considerations in therapy and assistance animals. In: Fine AH, editor. *Handbook on Animal-Assisted Therapy: Theoretical Foundations and Guidelines for Practice, 3rd Edn*. San Diego, CA: Elsevier Academic Press (2010). p. 26–200. doi: 10.1016/B978-0-12-381453-1.10023-6
17. Pastore C, Pirrone F, Balzarotti F, Faustini M, Pierantoni L, Albertini M. Evaluation of physiological and behavioral stress-dependent parameters in agility dogs. *J Vet Behav*. (2011) 6:188–94. doi: 10.1016/j.jveb.2011.01.001
18. McCullough A, Jenkins M, Ruehrdanz A, Gilmer MJ, Olson J, Pawar A, et al. Physiological and behavioral effects of animal-assisted interventions on therapy dogs in pediatric oncology settings. *Appl Anim Behav Sci*. (2018) 200:86–95. doi: 10.1016/j.applanim.2017.11.014

19. Melco AL, Goldman L, Fine AH, Peralta JM. Investigation of physiological and behavioral responses in dogs participating in animal-assisted therapy with children diagnosed with attention-deficit hyperactivity disorder. *J Appl Anim Welf Sci*. (2020) 23:10–28. doi: 10.1080/10888705.2018.1536979
20. Palestrini C, Calcaterra V, Cannas S, Talamonti Z, Papotti F, Buttram D, et al. Stress level evaluation in a dog during animal-assisted therapy in pediatric surgery. *J Vet Behav*. (2016) 17:44–9. doi: 10.1016/j.jveb.2016.09.003
21. SchoÈberl I, Wedl M, Bauer B, Day J, Möstl E, Kotrschal K. Effects of owner-dog relationship and owner personality on cortisol modulation in human dog dyads. *Anthrozoo*. (2012) 25:199–214. doi: 10.2752/175303712X13316289505422
22. SchoÈberl I, Beetz A, Solomon J, Wedl M, Gee N, Kotrschal K. Social factors influencing cortisol modulation in dogs during a strange situation procedure. *J Vet Behav*. (2016) 11:77–85. doi: 10.1016/j.jveb.2015.09.007
23. Hellhammer DH, Wust S, Kudielka BM. Salivary cortisol as a biomarker in stress research. *Psychoneuroendocrinolo*. (2008) 34:163–71. doi: 10.1016/j.psyneuen.2008.10.026
24. Fries E, Dettenborn L, Kirschbaum C. The cortisol awakening response (CAR): facts and future directions. *Int J Psychophysiol*. (2009) 72:63.e73. doi: 10.1016/j.ijpsycho.2008.03.014
25. Salimetrics. *Collection Method: SalivaBio Children's Swab (SCS)*. (2018). Available online at: https://salimetrics.com/wp-content/uploads/2018/02/children-swab-saliva-collection-instructions.pdf (accessed January 16, 2018).
26. Cohen S, Williamson G. Perceived stress in a probability sample of the United States. In: Spacapan S, Oskamp S, editors. *The Social Psychology of Health*. Newbury Park, CA: Sage (1988). p. 31–67.
27. Beetz A, Uvnas-Moberg K, Juilus H, Kotrchal K. Psychosocial and psychophysiological effects of human-animal interactions: the possible role of oxytocin. *Front Psychol*. (2012) 3:234. doi: 10.3389/fpsyg.2012.00234
28. Handlin L, Hydbring-Sandberg E, Nilsson A, Ejdebäck M, Jansson A, Uvnäs-Moberg K. Short-term interaction between dogs and their owners: effects on oxytocin, cortisol, insulin and heart rate—an exploratory study. *Anthrozoös*. (2011) 24:301–15. doi: 10.2752/175303711X13045914865385
29. Olff M, Frijling JL, Kubzansky LD, Bradley B, Ellenbogen MA, Cardoso C, et al. The role of oxytocin in social bonding, stress regulation, and mental health: an update on the moderating effects of context and inter individual differences. *Psychoneuroendocrinolo*. (2013) 38:1883–94. doi: 10.1016/j.psyneuen.2013.06.019
30. Rogers LJ. Relevance of brain and behavioural lateralization to animal welfare. *Appl Anim Behav Sci*. (2010) 127:1–11. doi: 10.1016/j.applanim.2010.06.008
31. Varga B, Gergely A, Galambos A, Kis A. Heart rate and heart rate variability during sleep in family dogs (*Canis familiaris*). moderate effect of pre-sleep emotions. *Animal*. (2018) 8:107. doi: 10.3390/ani8070107
32. Hennessy MB, Voith VL, Mazzei SJ, Buttram J, Miller DD, Linden F. Behavior and cortisol levels of dogs in public animal shelter, and an exploration of the ability of these measures to predict problem behavior after adoption. *Appl Anim Behav Sci*. (2001) 73:217–33. doi: 10.1016/S0168-1591(01)00139-3
33. Voith VL, Borchelt PL. Fears and phobias in companion animals. In: Voith VL, Borchelt PL, editors. *Readings in Companion Animal Behavior*. Trenton, NJ: Veterinary Learning Systems (1996). *p*. 140–52.
34. Beerda B, Schilder MB, van Hoff JA, de Vries HW, Mol JA. Behavioral and hormonal indicators of enduring environmental stress in dogs. *Anim Welfare*. (2000) 9:49–62. Available online at: https://www.researchgate.net/publication/40151127_Behavioural_and_hormonal_indicators_of_enduring_environmental_stress_in_dogs
35. Shiverdecker MD, Schiml PA, Hennessy MB. Human interaction moderates plasma cortisol and behavioral responses of dogs to shelter housing. *Physiol Behav*. (2013) 109:75–9. doi: 10.1016/j.physbeh.2012.12.002
36. Riemer S, Assis L, Pike TW, Mills DS. Dynamic changes in ear temperature in relation to separation distress in dogs. *Physiol Behav*. (2016) 167:86–91. doi: 10.1016/j.physbeh.2016.09.002
37. de Carvalho IR, Nunes T, de Sousa L, Almeida V. The combined use of salivary cortisol concentrations, heart rate and respiratory rate for the welfare assessment of dogs involved in AAI programs. *J Vet Behav*. (2019) 36:26–33. doi: 10.1016/j.jveb.2019.10.011
38. Clark SD, Martin F, McGowan RTS, Smidt JM, Anderson R, et al. Physiological state of therapy dogs during animal-assisted activities in an outpatient setting. *Animals*. (2020) 10:819. doi: 10.3390/ani10050819

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.

frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 25 November 2020
doi: 10.3389/fvets.2020.606388

Tilapia On-Farm Welfare Assessment Protocol for Semi-intensive Production Systems

Ana Silvia Pedrazzani[1]*, *Murilo Henrique Quintiliano*[2], *Franciele Bolfe*[2], *Elaine Cristina de Oliveira Sans*[1] and *Carla Forte Maiolino Molento*[1]

[1] Animal Welfare Laboratory, Federal University of Paraná (LABEA), Curitiba, Brazil, [2] FAI Farms, Londrina, Brazil

The aim of this study was to develop and test a tilapia on-farm welfare assessment protocol, based on Brazilian semi-intensive production systems. The study included two mains steps: the elaboration of tilapia welfare protocol and its on-field feasibility test. The protocol, including the potential indicators organized into health, environmental, nutritional, and behavioral categories, was tested on three farms. Skin, eyes, gills, jaws, fins, and vertebral spine were individually examined in 139 individual tilapias. Water physicochemical parameters and production system were considered. The overall nutritional status of individuals was assessed through body condition factor, feed conversion ratio, feed crude protein ratio, and feed ingestion behavior. During massive capture, signals of stress, level of crowding, and duration of air exposure were registered. Time required for loss of consciousness was evaluated by clinical reflexes and other behaviors during slaughter. Eye, jaw, and gill scores were different across farms (Kruskal-Wallis test, $p = 0.011$; 0.015; 0.043, respectively), showing good discrimination power. Critical welfare points were extremely low dissolved oxygen in water, fin and skin lesions, prolonged air exposure during pre-slaughter handling and non-humane slaughter techniques, as decapitation or asphyxia. The protocol presents practical viability and it is an initial step for the development of a tilapia welfare strategy, where the prioritization of critical welfare points, implementation of corrective actions and monitoring of the results is part of a permanent welfare management system.

Keywords: behavior, capture, fish, health, management, slaughter

OPEN ACCESS

Edited by:
Edward Narayan, The University of Queensland, Australia

Reviewed by:
Maurício Laterça Martins, Federal University of Santa Catarina, Brazil
Ana C. Strappini, Austral University of Chile, Chile

****Correspondence:***
Ana Silvia Pedrazzani
anasilviap@ufpr.br

Specialty section:
This article was submitted to Animal Behavior and Welfare, a section of the journal Frontiers in Veterinary Science

Received: *14 September 2020*
Accepted: *02 November 2020*
Published: *25 November 2020*

Citation:
Pedrazzani AS, Quintiliano MH, Bolfe F, Sans ECO and Molento CFM (2020) Tilapia On-Farm Welfare Assessment Protocol for Semi-intensive Production Systems. Front. Vet. Sci. 7:606388. doi: 10.3389/fvets.2020.606388

INTRODUCTION

In the last 20 years many studies regarding anatomic, physiologic, behavioral, and pharmacological aspects produced evidences that fish experience feelings such as pain and fear, in similar ways to other vertebrates (1, 2), as summarized in a text-book by Braithwaite (3). As evidences of fish sentience gain additional prominence (4), the concern about animal welfare by the society show parallel increases, affecting the consumer market and aquaculture regulations (5). This is so because, if fish are capable of suffering, then their welfare must be protected, within the same rationale employed for other vertebrate animals. In this context, the welfare of Nile tilapia *Oreochromis niloticus,* the most produced fish species in Brazil, may be considered a primary goal, as the number of individual animals involved is a criterium for priority in terms of animal welfare attention (6). Thus, there is an urgent need for new technologies, procedures, and strategies to detect and avoid or at least attenuate tilapia stress in all production stages and systems, so that their welfare may be improved.

In fish farming, welfare is likely to be compromised by routine management, causing stress due to the introduction of foreign objects into the water, the removal of the animal from the aquatic environment for individual restraint or underwater crowding (7). For example, the physical manipulation required for fish classification and biometrics management causes physical and psychological stress (8). In addition, water quality and associated environmental factors are areas of great attention by fish stress researchers, with water dissolved oxygen and carbon, pH and temperature and light regimes as the most critical environmental factors for maintaining fish homeostasis (9). Furthermore, there is a strong association between water quality and stocking density in fish farming systems, which is another important critical welfare point in aquaculture. When high density is associated with insufficient water renewal, the aversive effects are additive or multiplicative (1). In addition to deteriorating water quality, high density may increase aggression, lesions and disease. Under these conditions, parasitic infestations tend to thrive, generating high mortality rates (10). Additionally, restricted swimming space may also be detrimental to welfare (11). The optimal group size depends on the species behavioral characteristics and its tendency to form clusters or territorialism. Bhujel (12) suggests the optimal density of 5 animals/m^2 for tilapia, because lower densities promote aggression between males during breeding. Barcellos et al. (13) observed that tilapia fingerlings had detrimental effects from social stress including higher cortisol levels when maintained in a stock density of 10 fish/100 L^{-1}, when compared to 5 fish/100 L^{-1} and lower densities. As an additional complicating factor, the high density may influence food consumption, with dominant fish tending to eat more than others. A nutritionally balanced diet is critical to maintaining normal organic functioning and fish resistance to disease. Although periods of food deprivation may have an attenuated impact on fish balance due to ectothermia, consideration of their motivation to feed is essential in preserving welfare. Consequences of prolonged food deprivation may include aggression, erosion of the dorsal fin due to cannibalism and weight loss (14). When confined in high densities, some fish may suffer from fasting, again with potential additive adverse effects (10). Other critical welfare point is pre-slaughter and slaughter management. During massive capture, a procedure which before the recognition of fish sentience used to be named as harvest, and the transportation to the slaughterhouse, fish may suffer from multiple injuries due to overcrowding, resulting in excessive mucus production, loss of scales, damage to gill epithelium, muscle bruising and bone fractures, and extreme environmental conditions such as air and luminosity exposure as well as water quality shock. During slaughter, for a method to be considered humane it must immediately induce insensibility and be free from fear and pain (15–17). In Brazil, thermal narcosis employing ice slurry and the simplest air exposition causing asphyxia are the most common methods for tilapia slaughtering. However, these methods are not considered humane, due to the intense suffering and fear that fish experience for long periods before dying (15, 18–20).

Because of the number and diversity of welfare critical points, assessing the degree of farmed animal welfare requires the development of diagnostic techniques which are capable of considering an array of indicators. These can be classified as direct or animal-based indicators, when they are measured in the animals, or indirect or resource-based indicators, when they are measured in the environment in which the animal is inserted or relate to the management imposed on the animal (12). In addition, for the most common terrestrial farm animals, the Five Freedoms (21) are organized in robust welfare assessment protocols, such as the Welfare Quality (22) and the AWIN (23–25), which provide an organized list of specific welfare indicators to be measured and a first attempt of final integration into an overall animal welfare level. Fish-oriented scientific literature has also been building, and reported fish critical welfare points include indicators associated with feeding, water quality, sampling, capture, and slaughter (5, 7, 26). Additionally, interactions between different indicators may also present important effects on the welfare of fish.

Overall, the effectiveness of a welfare protocol depends on its validity, reliability, and feasibility. In other words, the protocol needs to be validated by expert judgment (27, 28), repeatedly achieve the same results by the same or different observers after adequate training, and be consistent in terms of required time and across different farm conditions (29). Although there are some protocols for some fish species as salmon and trout (30–33), optimal values of indicators are species-specific and no tilapia welfare protocol assessment seems to be available in the literature. Therefore, the goal of this work is to develop a tilapia on-farm welfare assessment protocol, based on Brazilian semi-intensive systems of production, bringing fish welfare assessment efforts closer to the more robust literature on the welfare assessment of terrestrial farm animals.

MATERIALS AND METHODS

The study included two mains steps: the elaboration of tilapia welfare protocol and its on-field feasibility test. For the elaboration of the protocol, initially a list of potential indicators was prepared from a literature review. Then, 12 tilapia farms in South and Southeast of Brazil were visited, for studying the measurement feasibility of each selected indicator, in terms of time and human resources required. Thus, the tilapia welfare protocol was ready for the second main step, which was its on-field testing on three different farms, chosen from our contacts with the criterium of nutritional and mortality data availability. Field results were studied in terms of each indicator potential for contributing to the overall welfare assessment and to the discriminating power across different real-life conditions relevant to the target production system.

This project was approved by the Animal Use Ethics Committee of the Agricultural Campus (No. 083/2019), of the Federal University of Paraná, Brazil.

Elaboration of the Tilapia Welfare Assessment Protocol

The tilapia welfare assessment protocol was organized in four categories as per the literature on farm welfare assessment protocols for terrestrial animals (22): (a) health, (b) environment, (c) behavior, and (d) nutrition, considering the severity and duration of potential risks (**Table 1**). Health and environmental indicators were established based in salmon protocols and adapted for tilapia (**Tables 2**, **3**) (30, 31). Tilapia environmental and nutritional needs, as per the scientific literature, were used to adjust the scores criteria. Finally, behavioral indicators were incorporated to the protocol. Scores were set for all categories, with 1 representing the desirable scenario.

Ten health indicators were based on clinical examination of the eyes, jaws, operculum, skin, fins gills, and spine, the presence of ectoparasites, blood glucose, and mortality (**Table 2**). Environmental indicators included seven water physicochemical factors, stocking density, the presence of interspecific cohabitants, shading, and terrestrial predators (**Table 3**). For nutritional assessment the indicators were body condition factor, dietary crude protein level, conversion ratio, and feeding behavior (**Tables 4**, **5**). Condition factor (K) was defined as $K = (WL^{-3})100$ to estimate tilapia nutritional status, where W is the weight (g) and L is the length (cm) (12, 27). The K factor was classified according to the value obtained, with score 1 for K between 1.6 and 1.9; score 2 for 1.1–1.5 or 2.0–2.3; and score 3 for ≤1.0 or ≥2.4. Stocking density was also relevant for nutritional evaluation, which was classified as adequate or inadequate according to life stage recommendations (34). For all nutritional indicators, score 1 was the ideal scenario, 2, 3, and

TABLE 1 | Health, environmental, behavioral, and nutrition indicators for the assessment of farmed tilapia welfare, based on Stien et al. (31).

	Welfare indicator	Production stage		
		Growing/grow-out	Capture	Pre/slaughter
Health	Eyes, jaws, operculum	X	X	X
	Skin, fins, gills	X	X	X
	Spine	X	X	
	Ectoparasites	X		
	Blood glucose	X	X	X
	Mortality	X	X	
	Scales in water		X	X
	Consciousness			X
Environmental	Temperature, pH	X	X	X
	OD, NH_4, NH_3, NO_2	X	X	
	Transparency	X		
	Stocking density	X	X	X
	Shading	X	X	X
	Predators control	X	X	
	Interspecific	X		
	Air exposure		X	X
	Light exposure		X	X
Behavioral	Gulping air at surface	X	X	
	Respiratory frequency	X	X	X
	Swimming	X	X	X
	Distribution in tank	X	X	
	Body coloration	X	X	
	Social behavior	X		
	Foraging behavior	X		
	Response to light	X	X	X
	Response to air exposure		X	X
	Loss of consciousness			X
Nutritional	Amount of feed provided	X		
	Crude protein (CP)	X		
	Feed conversion ratio	X		
	Condition factor (k)	X		
	Feeding handling	X		
	Fasting period	X	X	
	Depuration period			X

TABLE 2 | Health indicators based on Stien et al. (31), scores and descriptors or reference values adapted for on-farm tilapia welfare evaluation.

Indicators	Score	Descriptors or reference values
Eyes	1	Apparently functional and healthy
	2	Hemorrhage, exophthalmos, traumatic injury; Unilateral
	3	Hemorrhage, exophthalmos, traumatic injury; Bilateral
	4	Bilateral cataract, chronic condition, impaired vision
Jaws	1	Normal aspect, healthy
	2	Light superior or inferior deformity (esthetics)
	3	Moderate superior or inferior deformity (affecting feeding)
Operculum	1	Normal aspect, healthy
	2	Partially covering the gills (≥75% covered)
	3	Partially covering the gills (<75%)
	4	Unilateral or bilateral absence
Skin	1	Normal aspect, healthy
	2	Scar tissue, scale loss, ulcers or superficial injuries <1 cm^2
	3	Ulcers or superficial injuries >1 cm^2, redness, light necrosis
	4	Severe necrosis, darkening, bleeding, inflammation
Fins	1	Normal, healthy appearance
	2	Scarred or slightly necrotic tissue
	3	Moderate injury or necrosis (thickening/splitting)
	4	Severe necrosis, bleeding, inflammation, exposure of the rays
Gills	1	Normal aspect, healthy
	2	Light injury or necrosis, thickening or splitting
	3	Moderate injury or necrosis, thickening or splitting
	4	Severe necrosis, bleeding, inflammation, pallor, or darkening
Spine	1	Normal structure
	2	Lordosis or scoliosis, normal weight
	3	Lordosis or scoliosis, weight loss
Ectoparasites	1	No infestation
	2	Moderate infestation (≤5 parasites)
	3	Intense infestation (>5 parasites)
Blood glucose (mg/dL)	1	30–59
	2	60–80
	3	81–120
	4	<30; >120
Mortality (%)	1	≤10
	2	≤25
	3	≤50
	4	≥75

TABLE 3 | Environmental indicators based on Stien et al. (31), scores and descriptors or reference values adapted for on-farm tilapia welfare evaluation.

Indicators	Score	Descriptors or reference values
Temperature (°C)	1	25–32
	2	20–24
	3	33–37
	4	≤19–≥38
pH	1	6.0–8.5
	2	5.5–5.9 or 8.6–8.9
	3	8.9–10.0
	4	≤5.5 or ≥10.0
Transparency (cm)	1	25–40
	2	41–65
	3	<25 or >65
Oxigen saturation (%)	1	70–95
	2	50–69
	3	30–49
	4	<30 or >95
Non-ionized ammonia (NH_3; mg/L)	1	0.000–0.050
	2	0.050–0.100
	3	≥0.100
Nitrite (NO_2; mg/L)	1	0.00–0.50
	2	0.50–1.00
	3	≥1.00
Alkalinity (mg/L of $CaCO_3$)	1	30–100
	2	20–30 or 100–200
	3	<20 or >200
Shading (%)	1	20 a 30
	2	31 a 40
	3	<20 or >40
Predators	1	Absence
	2	Controlled presence
	3	Uncontrolled presence
Interspecific inhabitants	1	Absence
	2	Controlled presence
	3	Uncontrolled presence
Stocking density	1	Ideal to 10% overpopulation
	2	10–20% overpopulation
	3	> 20% overpopulation

4 being off in 10, 20, and more than 20% of optimal values, respectively. Feeding behavior was classified as appropriate if fish consumed the feed within 3–5 min. The swimming behavior and the level of fish body air exposure during capture, and the time for the loss of consciousness after stunning or slaughter procedures were also included in the protocol (**Table 5**). The indicators for the evaluation of tilapia consciousness included the clinical reflexes: opercular rate (OR), vestibulo-ocular reflex (VOR), equilibrium (EQ) and the tail-grab-reflex.

On-Field Feasibility Test of the Tilapia Welfare Assessment Protocol

The welfare was assessed using potential indicators during summer (December, 2019–March, 2020) at two different grow-out excavated pond farms (A and B) located in Joinville, Santa Catarina, and one grow-out tilapia cage facility (farm C) in Guaíra, São Paulo, respectively, in Southern and Southeastern Brazil. Farm A was composed by six rectangular excavated ponds built on flat ground, and the assessment of water quality and massive capture was performed in two of them, with areas of 3.385 and 5.050 m^2. The circulation system was maintained by a diversion canal and the water flow was controlled through a water intake and drained by a drainage canal. On farm B there was one 2.115 m^2 excavated pond, that was supplied from the water-table by seepage into the pond. On farm C there were six excavated

TABLE 4 | Different densities considered ideal according to the association between raising system and tilapia feed conversion ratio (FCR) and diet crude protein rate (CP), adapted from RSPCA (33).

Raising system	Weight (g)	Age (days)	Stock density (fish/m²)		FCR	CP (%)
			No aeration or renew	Aeration or renew		
Excavated pond	1–30	40–80	20–30	40–50	0.8–1.0	36-40
	30–200	80–120	4–5	6–10	1.2–1.3	28–32
	200–1,000	>120	0.8–1.2	2–3	1.4–1.6	28–32
Cage	1–30	40–90	1,200–1,500		0.8–1.0	40
	30–200	90–120	450–500		1.2–1.4	32
	200–1,000	>120	100–150		1.6–2.0	32

TABLE 5 | Scores used to classify feeding, capture, and slaughter indicators and respected characteristics for on-farm tilapia welfare assessment, based in Noble et al. (26).

Management	Score	Criteria
Feeding	1	Apprehension of all food in 180–300 s
	2	Apprehension of all food in 120–180 s
	3	Apprehension of all food in ≤120 s
	4	No apprehension of all food or ≥360 s
Capture	1	Normal swimming, no or low dorsal fins or body parts on surface
	2	Excited swimming behavior, >20 dorsal fins or low body parts on surface
	3	Swimming in different directions or decreasing activity, fish stuck against net
	4	Many fish floating on side, explosion of body to air, exhaustion
Slaughter	1	Instantaneous loss of VER, BO, EQ, and TGR
	2	Total loss of VER and BO in ≤10 s, instantaneous loss of EQ and TGR
	3	Total loss of VER and BO in ≤100 s, instantaneous loss of EQ and TGR
	4	Total loss of VER and BO in ≤1,000 s, progressive loss of EQ and TGR

ponds as well as fish cages placed inside them or directly in a river channel. For farm C, the study was performed in two different excavated ponds of an area of ∼10.000 m², each containing 28 and 42 floating steel fish cages of 4.8 m³ (2.0 × 2.0 × 1.2 m) with a 20 mm steel galvanized mesh covered with PVC and four floaters: The water was renewed both indirectly by gravity and pumping through a diversion canal from a reservoir. There was were an individual inlet and an individual an outlet for each pond.

Indicators were measured always by the same researcher, during each farm routine schedule and with minimum interference to the daily management and procedures. The samples sizes were defined according to the farm dynamics, assessing the maximum number of individuals without disturbing farm routine. On farms A and B, it was possible to access a larger sample, derived from massive capture. On the other hand, due to the complexity of tests performed on farm C, where slaughter was performed, the number of individuals was reduced, avoiding disturbances to the slaughter processing line. In total, 139 tilapias were physically scored: 72, 40, and 27 animals on farms A, B, and C, respectively. Fish were removed from water, placed in a desk covered with soft material, identified using numbered waterproof cards, which were placed beside the animal for bilateral photographic registration. Sequentially, the tilapias were weighed, measured, and physically assessed according to the scores set in the protocol. On farms A and C, fish were in the end of final grow-out stage phase and thus destinated to slaughter after individual scoring. On farm B, as tilapia were in the intermediate grow-out phase, they were returned to the pond of origin.

All the environmental indicators were assessed minutes before removing fish from water for massive capture (farms A and B) or for biometrics (farm B). The physicochemical indicators were measured directly in the water by insertion of an equipment for multi-parameter measurement (AK 87, Akso, Brazil). The depth of 30 cm was standardized for measuring water temperature, pH and dissolved oxygen (DO). Water samples were collected for conducting the colorimetric test of total ammonia (NH_4), nitrite (NO_2) and alkalinity (Acquacombo TD 1555, Alfakit, Brazil), tested immediately after collection. Non-ionized ammonia (NH_3) was estimated using a specific formula considering water temperature and pH (35).

Feeding behavior was measured considering the time taken for the food to be fully consumed by the animals. Production indices were collected via interview with farm staff, mostly regarding mortality rate, stocking density and daily amount of feed provided, to calculate feed conversion rates.

The welfare score of massive capture was assessed once on farms A and C, according to the capture method adopted in each place, which directly influenced the length of the light and air exposure periods, as well as the level of crowding (**Table 4**). On farm B, the animals studied were those captured for farm routine fish biometric procedure, using a fishing net. According to the husbandry normally adopted on farm C, after the capture fish were weighed in groups of 20 animals and individual weight was estimated by the group mean. For this reason, it was not possible to measure the fish length, so the K factor was not calculated for the animals studied on farm C.

After massive capture, the slaughter score was assessed in a slaughterhouse attached to farm C. Health exam and blood glucose analysis were made before fish slaughter. Blood was punctured from caudal vein using a 25 × 7 mm needle coupled to a 3 mL syringe. Blood was then transferred to a glucometer strip for instantaneous glucose measurement (AccuCheck Active, Roche, Brazil). In sequence, the slaughter technique was assessed, based on the registration of the time needed for the loss of consciousness (36). The OR is the simplest way to estimate the respiration rate, by counting the opercular movement, which is inversely proportional to the level of consciousness. The VOR is measured by the visual evoked response (VER) or "eye roll," that is the eye movement when fish body is rolled from side to side through the vertical axis (37). The EQ was evaluated by fish position and its swimming capacity when placed into the water. TGR is the grabbing of the animal's tail to verify whether the fish attempts to escape (26), being an effective way to evaluate the capability of fish to interact with the external environment (15).

Statistical Analysis

Descriptive analysis was used to estimate the welfare scores obtained during the field evaluations. The normality of data was tested by Shapiro-Wilk test ($p < 0.05$), using Statistica Statsoft 10.0. T-test was applied to compare farms A and B in relation to factor K results, because it was the only variable with normal distribution ($p = 0.054$). The non-parametric results obtained for health indicators on farms A, B, and C were compared by Kruskal-Wallis, aiming to detect some uniformity of results in relation to the critical welfare points observed in different facilities. Correlation analyses ($p < 0.05$) were conducted aiming at an enhanced comprehension of the relationship between environment conditions and health results.

RESULTS

Health Indicators

Health indicator results are summarized in **Table 6**. When comparing scores from different farms, differences were observed for eyes, jaws and gills (**Figure 1**). On Farm A, few animals presented damage in different degrees of severity in the eyes, being opacity, unilateral, and bilateral hemorrhage. From the total of 72 animals examined in that farm, 10 tilapia (13.8%) presented damaged jaws and two had unilateral partial loss of operculum. Spine seemed to be healthy in most animals, with just one case of scoliosis. Also, when evaluating ectoparasites, just one parasitic copepod commonly called anchor worm *(Lernaea* sp.), attached to the skin was detected. Light gill lesions without color alteration were observed in 33 fish (45.8%); five animals had moderate lamella fusion and excessive mucus production. From that, 36 animals (50.0%) showed splitting of caudal fin, and seven of these had additional necrotic dorsal fins. The main alteration observed in skin was subcutaneous hemorrhage, usually located in fish face between snout and operculum. On farm B eye damages, vertebral deformations or ectoparasites were not registered. However, from the 40 tilapia, 11 animals (27.5%) showed jaw lesions and 14 (35.0%) presented red spots on other areas of face skin. Gill splitting and excessive mucus production were observed in 10 fish (25.0%). Caudal fins of 19 animals (47.5%) were splitted and one tilapia presented dorsal fins light necrosis (2.5%). On Farm C, one tilapia (3.7%) from the total of 27, presented lesions on the jaw and operculum. When examining the eyes, two animals (7.4%) had unilateral hemorrhagic lesion and three animals presented bilateral exophthalmos. Gills of nine animals (33.3%) showed partial damage (lamella fusion) and two animals (7.4%) exhibited severe damage, including two positives for parasites with macroscopic signals suggestive of the monogenoid *Dactylogyrus* sp. Future work may include microscopic diagnosis of eventual parasite species. As for the fins, 21 animals (77.8%) presented light erosion of one or more fins, and one of them presented severe necrosis (3.7%). Coherent with clinical findings for the fins, when examining body skin, different alterations were observed, as lack of scales ($n = 7$; 25.9%), ulcerations ($n = 2$; 7.4%), necrosis ($n = 2$; 7.4%), and body skin darkening ($n = 1$; 3.7%). No skeletal deformities were detected. Blood glucose level, measured on farm C immediately before slaughter, was 86.44 ± 13.75 mg/dL, distributed in scores 1 (5.6%), 2 (27.8%), and 3 (66.7%). According to farm staff data, the mortality along the production cycle on farms A and C was 5 and 10%, respectively, and both were classified as 1; on farm B reported mortality was 15%, thus equivalent to score 2.

TABLE 6 | Health scores (%) and p-values in three different tilapia grow-out farms, data collected from January to March 2020 in South and Southeast Brazil; comparison amongst farms with Kruskal-Wallis test for all indicators.

Health indicator	Farm A ($n = 72$)				Farm B ($n = 40$)				Farm C ($n = 27$)				
	1	*2*	*3*	*4*	*1*	*2*	*3*	*4*	*1*	*2*	*3*	*4*	*p*
Eyes	95.8	1.4	1.4	1.4	100	0.0	0.0	0.0	81.5	7.4	11.1	0.0	0.011
Jaws	86.1	13.9	0.0	–	72.5	27.5	0.0	–	96.3	3.7	0.0	–	0.015
Operculum	97.2	2.8	0.0	0.0	100	0.0	0.0		96.3	3.7	0.0	0.0	0.475
Gills	48.6	45.8	5.6	0.0	75.0	25.0	0.0	0.0	59.3	33.3	7.4	0.0	0.040
Skin	52.8	36.1	11.1	0.0	65.0	35.0	0.0	0.0	55.6	25.9	14.8	3.7	0.311
Fins	40.3	50.0	9.7	0.0	50.0	47.5	2.5	–	18.5	77.8	3.70	–	0.056
Spine	98.6	1.4	0.0	–	100	0.0	0.0	–	100	0.0	0.0	–	0.543
Ectoparasite (*Lernaea* sp.)	98.6	1.4	0.0	–	0.0	0.0	0.0	–	100	0.0	0.0	–	0.543

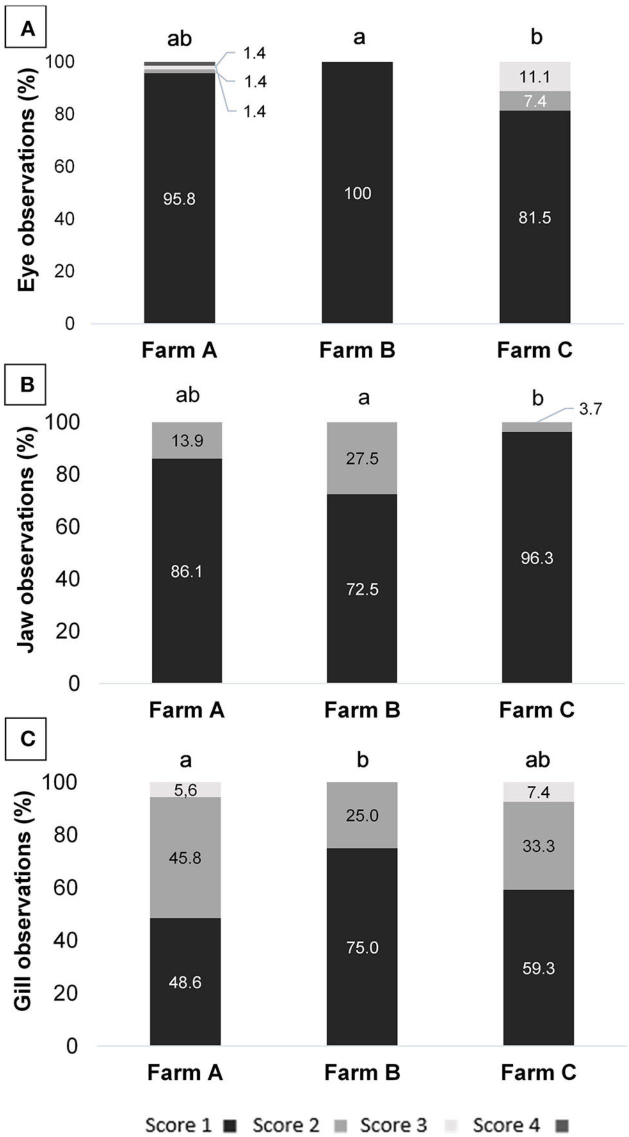


FIGURE 1 | (A) Eye, **(B)** jaw, and **(C)** gill scores assessment of tilapia from three different fish farms; different letters indicate significance obtained by Kruskal-Wallis test ($p = 0.011$; 0.015; and 0.043, respectively).

Environmental Indicators

The excavated pond areas were 3.385 and 5.050 m^2 on farms A and B, respectively. On farm C, an excavated pond accommodated 28 tilapia cages measuring 4.8 m^3 (2.0 × 2.0 × 1.20 m), totaling 134.4 m^3 of area available to the animals. Even though stocking density was not considered elevated, farms A and B presented critical levels of DO, of 1.6 and 1.7 mg/L respectively (**Table 7**). On farm A the pH was 6.5, the limit of water acidification considered adequate for tilapia, probably related to low alkalinity and high transparency of water. In all scenarios, few concerns were observed about the pond's external environment. In this way, unsatisfactory scores were observed in relation to the absence of shading. Also, it was detected uncontrolled presence of terrestrial predators, mainly birds in farm A and B. In farm C fish were protected by a 25 mm galvanized steel screen cover, avoiding predation. Despite that, many birds were present above it and in land near cages. The maintenance of tilapia as a unique species was adopted in all evaluated systems, so aquatic predators or competitors were not detected.

Nutritional Indicators

In all scenarios, fish were self-fed twice daily with commercial extruded pellets. Even for fish maintained in the pond for an extended period on farm A, according to stocking density informed by staff and the amount of feed provided, the FCR calculated was considered adequate (**Table 8**). In farm B, the crude protein ratio was 6% higher than the suggested for weight and age criteria, being classified as welfare score 2. K factor means for farms A and B were 1.52 ± 0.50 and 2.45 ± 0.50 ($p = 0.000$), scored as 2 and 3, respectively.

Farms A and B adopted similar protocols for delivering the feed, where handling was made manually from just one pond margin side. The superficial swimming efforts of fish in the direction of the feed pellets were noticeable. Farm A had a larger swimming area, being difficult for all the fish to achieve the pellets during feeding time. In both situations, the consumption occurred within 5 min. Feed management on farm C cages was carried quickly by canoe, taking around 20 min to cover all the 28 cages. The feed intake was quite fast, ∼1 min for total feed consumption. This may be an indicative of underestimation of the amount of feed to offer, especially for days of intense heat, when the metabolism of the fish is accelerated. However, an excess of feed was observed in cages near the margins, probably due to low consumption associated with the stress of massive capture, as fish caught and considered underweight, called rejects, were reallocated into these same cages.

Behavioral Indicators

The massive capture occurred between 7 h 00 and 7 h 30 a.m. in all scenarios, which likely provides less stress than would be caused by the stronger light and heat of the most advanced hours of the day. On farm A, fish were caught by five men pulling a fishing net from inside the pond. The procedure lasted more than 3 h. After collection, animals remained overcrowded and stuck in the fishing net, being exposed to sun light for around 30 min, until being gradually removed from water manually. Most fins and body parts could be observed over the water column and exhaustion was evident through the intensive swimming as tentative to escape. Fish were then placed in dry plastic boxes (35 animals/box) and placed in 1,000 L transport boxes located over a fish transport truck, containing water and ice. On farm C the procedure was faster, as the massive capture, including the cage displacement to the handling deck and lifting, the fish capture by fishing net (∼10 fish/catch) and weighing lasted 20 min. However, some critical welfare points in relation to this management were observed. Fish that were not sold to customers immediately remained in same cage until slaughter time by the afternoon. In the meantime, the water column was extremely reduced due to the tank lifting. Fish presented agitated swimming

TABLE 7 | Environmental indicators values and scores of three different tilapia grow-out farms, data collected from January to March 2020 in South and Southeast Brazil.

Environmental indicator	Farm A		Farm B		Farm C	
	Value	Score	Value	Score	Value	Score
Temperature (°C)	28.0	1	26.5	1	29.8	1
pH	6.5	1	7.0	1	7.5	1
Transparency (cm)	31.0	2	28.0	1	22.0	1
DO (%)	19.0	4	21.0	4	50.8	2
NH_4 (mg/L)	0.060	–	1.821	–	0.815	–
NH_3 (mg/L)	0.001	1	0.011	1	0.019	1
NO_2 (mg/L)	0.000	1	0.000	1	0.000	1
Alkalinity (mg/L)	20.0	2	30.0	1	40.0	1
Shading (%)	0.0	3	0.0	3	0.0	3
Predators	UP[a]	3	UP[a]	3	CP[b]	2
Inhabitants	NI[c]	1	NI[c]	1	NI[c]	1
Density ([d]fish/m^2; [e]fish/m^3)	1.3[d]	1	3.6[d]	1	70.0[e]	1

[a]Uncontrolled presence; [b]Controlled presence; [c]Non interspecific inhabitants. [d]fish/m2; [e]fish/m3.

TABLE 8 | Nutritional indicators and related information for three different tilapia grow-out farms data collected from January to March 2020 in South and Southeast regions of Brazil.

Technical information and nutritional indicators	Excavated pond		Cage
	Farm A	Farm B	Farm C
Fish weight (g) (mean ± SD)	737.9 ± 132.6	274.2 ± 39.26	1080.5 ± 229.8
Fish age (days)	418	118	400
Density ([a]fish/m^2; [b]fish/m^3)	1.30[a]	3.61[a]	70.00[b]
Crude protein ratio (CP) (%); score	32/1	38/2	32/1
Feed conversion ratio (FCR); score	1.45/1	1.54/1	2.00/1
K factor (mean ±SD); score	1.52 ± 0.50/2	2.45 ± 0.50/3	–
Feeding (min); score	5 min/1	5 min/1	1 min/3

[a]fish/m2; [b]fish/m3.

and some parts of body were exposed to air and luminosity. In farm B the capture using fishing net was performed four times, aiming to collect enough animals to conduct biometrics. Fish were placed in 10 L bucket (1 fish L^{-1}) containing pond water for ~15 min. Even if its duration was faster than the massive capture observed in other scenarios, animals from farm B showed attempts to escape and acceleration of opercular beating. Considering the period of capture, handling, and its consequent air exposure and crowding, farms A and B were classified as score 2 and Farm C was classified as score 4. As for pre-slaughter and slaughter, on farm A fish were placed in crowded transport boxes containing water and ice until transport to the slaughterhouse. No control of temperature, DO or stunning effectiveness was observed. In the first minutes after allocation in the transport container, animals presented agitated swimming and frequent escape behavior, and at the end of the capture procedure the first animals to be submitted to the ice slurry were apparently dead. On farm C, 10% of animals were sold directly to the local market. In this case, the tilapias were placed in raffia bags without any procedure for slaughtering or stunning them, and consequently probably died from asphyxiation on the way to the reseller. The animals considered too small were thrown alive to be consumed by birds, constantly present around the tanks likely conditioned to that practice. Remaining fish, on farm C, were placed in 500 L tanks containing water and arranged in a truck; after arrival at the slaughterhouse complex, the animals were transferred to another tank. This slaughterhouse tank had a recirculation pump; however, there was no filtering system or temperature control device. Animals that were slaughtered lastly showed signs of physical exhaustion, apathy and remained practically outside of the water because the tank was drained before the slaughter procedure was complete. On the processing table, tilapias were decapitated with a knife and fileted. It was noticed that after decapitation, some organs as the heart and pectora fins stayed connected to the fish head. This allowed for the presence of movements which characterize consciousness in most severed heads. From 10 animals that were evaluated, severed heads presented frequent OR, six showed attempts to swim when reconditioned into water and three showed mouth regular opening movements. The average time for loss of OR and VER were 257.36 ± 121.42 and 301.87 ± 120.16 s, respectively. However, no fish showed a reaction to the pain stimulus applied in the lips after decapitation. Due to the prolonged suffering that animals were exposed to in both scenarios, slaughter welfare score was classified as 4. In total, 26 welfare scores were measured on farms A and C, and 25 evaluated on farm B, including health, environment, nutrition and behavior indicators. For the comparison between farms analysis, all indicators scores were included. Despite the high frequency of score 4 on farm A, when considering the total added welfare scores, no significance between farms was found (**Figure 2**; $p = 0.435$). Results of analyses showed a weak correlation between gill score and weight (**Table 9**). The K factor was moderately correlated to DO and pH. No significance or very weak correlations were observed between health indicators (gills, fins, and skin) scores and environmental indicators.

DISCUSSION

To improve animal welfare in fish farming standardized protocols to assess welfare are needed (38). The main objective of this work was to develop a tilapia welfare assessment protocol and test its discrimination power to prioritize critical welfare points as well as to show differences amongst real life situations. The robustness of the protocol was possibly increased due to the observed variability across farms, as an extensive list of indicators was included, making the protocol more suitable for generalization to an array of realities. Even then, this is a preliminary protocol to which more indicators will likely be added in the future, as shown in the case of species in which animal welfare assessment has been studied for longer [see for example Souza et al. (39)].

The quality of the data reported by local staff, mainly regarding stocking density and mortality rate, is a relevant factor to consider, as the lack of accurate information can preclude the calculation of the adequate animal sample size for individual evaluation and, consequently, a precise welfare assessment. For example, the mortality considered here refers to the total percentage recorded over the grow-out cycle, i.e., historical mortality. Rates of mortality vary considerably depending on the production stage and farm management adequacy, with 20–71% mortality reported for tilapia, and is an important tool for the identification of critical welfare points identification (40). However, in addition to the lack of this indicator usually observed on small farms, a low mortality rate does not guarantee the absence of disease (26) or a high degree of welfare (5).

Results of health indicators showed important variability in gill, fin, and skin conditions. Gills are vital organs, and as they are exposed to the external environment, their changes are generally visible and can indicate various diseases such as

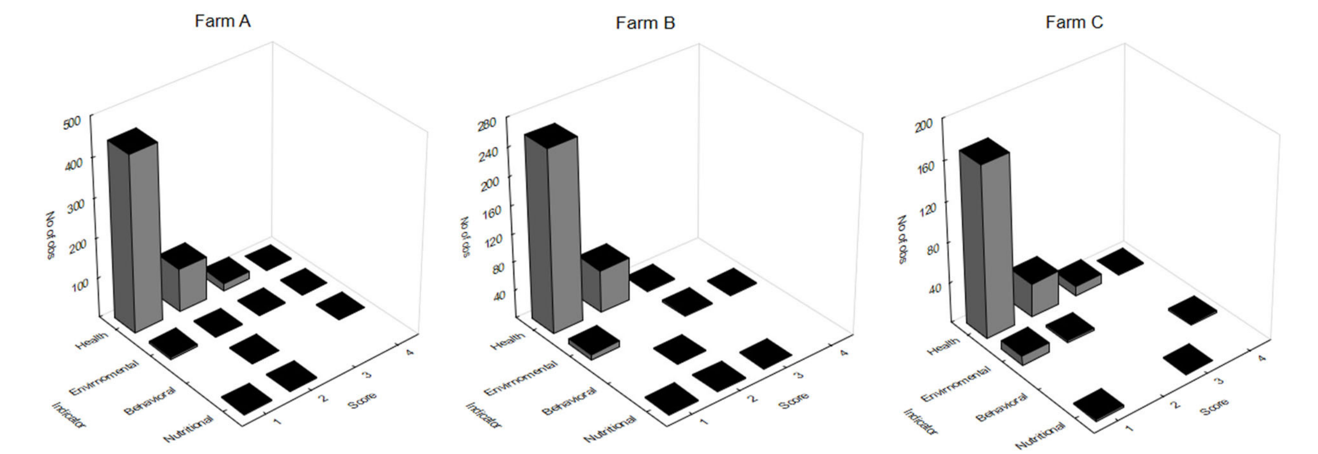


FIGURE 2 | Tilapia welfare scores obtained for health, environmental, nutritional, and behavioral indicators in three different farms **(A-C)**; no significance was found when comparing farms (Kruskal-Wallis; $p = 0.435$).

TABLE 9 | Correlation between different health, nutritional and environment indicators of tilapia welfare, showing *p*-value, coefficient of determination (r^2), tendency and correlation strength.

Variable	Gills	Fins	Skin	K factor
Weight	0.01; 0.07; (+) VW	0.09; 0.03; (+) NS	0.55; 0.00; (+) NS	0.00; 0.34; (–) WK
DO	0.47; 0.00; (–) NS	0.20; 0.01; (+) NS	0.06; 0.03; (+) NS	0.00; 0.41; (+) MD
Temperature	0.99; 0.00; (+) NS	0.42; 0.00; (+) NS	0.04; 0.03; (–) VW	0.00; 0.24; (+) WK
pH	0.13; 0.02; (–) NS	0.14; 0.02; (–) NS	0.16; 0.01; (–) NS	0.00; 0.51; (+) MD
NH_3	0.17; 0.01 (–) NS	0.89; 0.00, (+) NS	0.23; 0.01; (–) NS	0.17; 0.02; (+) NS

Correlation strenght based on r^2 range: (VW) Very Weak: 0.00–0.19; (WK) Weak: 0.20–0.39; (MD) Moderate: 0.40–0.69; (ST) Strong: 0.70–0.89; Very Strong (VS): 0.89–100; (NS); Not significant ($p > 0.05$).

parasitosis, bacteriosis, and viral infections (30). Gill aspect is also considered an important indicator of water quality, and its alterations may reflect signals of inadequate pH or intoxication by high levels of ammonia, nitrite, or chlorine (41). In addition, when subjected to prolonged air exposure, gill lamellae collapse and adjacent filaments adhere, reducing the gas exchange area and causing hypoxia (42). This gill collapse or division was a common alteration observed during health evaluation in farms A and C, after massive handling, indicating this excessive air exposure during the procedure.

Similarly, fin alterations may be a signal of various diseases, inappropriate handling or cannibalism. The relationship between severity, frequency and type of fin damage and welfare is not well-understood (5, 26). However, fins are composed of hemispheric tubes containing blood vessels and nerve bundles of nociceptors; therefore, a fin lesion may be painful (43). Also, fin damage may have a detrimental effect upon growth and survival, and may potentially reduce swimming ability, affecting fish welfare (19). In this way, it is important to determine if the lesion is acute, being recently induced by the management itself, or chronic. Fin chronic erosion can occur as a primary or secondary consequence of bacterial or fungal disease (44). In farms A and B, it was possible to observe high prevalence of caudal fin splitting, operculum damage, and redness in the skin of the tilapia head, all characteristic lesions of capture using knotted nets (45–47). The tendency toward a significant result when comparing fin conditions amongst the three farms pointed to this influence of capture management, explaining the higher splitting occurrence in farms A and B. Differently, the injuries with higher prevalence and severity observed in farm C, are characteristic of bacterial necrosis. Erosions in fins can also be caused by turbine pumps and crowding during capture, feeding handling, water temperature and oxygen supersaturation, exposure to light and consequent sunburn (19, 26). The preventive measures include the use of knotless netting or vacuum pumping for fish transference, and the use of demand feeding technology, avoiding cannibalism (19, 26, 44). Specific and more detailed investigations of types of lesion and on which fins they usually occur are recommended for tilapia welfare evaluation, being an important tool for farmed fish welfare improvement.

The significant difference observed for eye scores between farms may relate to the sanitary condition of each property. As eye alterations are indicative of several pathologies, this indicator presented good discriminatory power, exposing the problem observed in farm C. Ocular disorders in fish are common and may occur as primary or secondary manifestations of a systemic disease (48). Exophthalmos is a clinical sign of important bacterial pathologies that affect tilapiculture, such as *Aeromonas hydrophila, Streptococcus agalactiae, Flavobacterium columnare* (49), and orthomyxo-like virus, the agent of tilapia lake virus disease (TiLV) (49–51). These viral and bacterial co-infections are common (50). Many of the skin damages observed during our assessment, such as hemorrhages, ulcers and body darkness, were also compatible with bacterial infections, mechanical injuries or stress caused by handling. The fish epidermis is a multifunctional organ with highly relevant physiological roles, including a cutaneous stress response that protect the organism against unfavorable conditions (52). If this barrier is physically lesioned, the organism is more susceptible to infections. Additionally, the skin is loaded with nociceptors, and thus every damage potentially painful, considering lesion frequency and severity. Overall, epidermal damage is easy to evaluate and an important welfare indicator as it indicates serious welfare concerns (44). It also revealed important prevalence, with scores 2 and 3 in all real scenarios studied.

Despite the significance amongst farms, jaw evaluation evidenced an acute welfare problem, as the lesions seemed to be resultant from the net crowding during massive capture (45). Another indicator that showed low prevalence was the presence of ectoparasites. However, parasitic diseases are among the most frequent problems in aquaculture and are frequently associated with inadequate water conditions and high densities (42). Tricodinids and monogenoidea are among the most important ectoparasites in tilapia (53–55), with high parasite specificity to the gills and skin during the warmest months, causing discomfort and death by asphyxia (53, 56). However, their definitive diagnosis is only possible through the microscopic analysis. As the intention is to develop an on-field protocol to be executed by farmers, alternatives to laboratory analyses are preferred (22, 31, 57). Bui et al. (58) proposed behavioral observation as an indicator of Atlantic salmon sea lice infestation, as the increase in flashing or jumping behavior are potential signals of higher parasite prevalence. In addition, the behavior may be assessed by farmers through standardized visual surveillance or through more advanced methods such as assessment of shoaling behavior from video recordings (58). As ectoparasites represent a critical fish welfare issue, affecting locomotion, competition skills, and foraging behavior (59), the development of specific behavioral measures for the assessment of responses to parasites in tilapia is relevant to the management of their welfare.

Blood glucose is an invasive measurement when dealing with tilapia, considering the air exposure, hemorrhages and skin lesions that may occur and carry on relevant risks for the future welfare of the animals, such as for instance the development of infections. In this work, fish were sampled immediately before slaughter due to these risks. In addition, the procedure is time consuming for handlers, requiring specific skills that may not be common for farm staff. Furthermore, a set of rapid, inexpensive and non-invasive screening methods is preferable as on-field welfare indicators (38). It is also fundamental to consider that glucose levels in this setting represent the intensity of the stress related to massive capture, transportation, and other pre-slaughter interventions, being less related to the stress levels during life. This indicator is more meaningful if compared with pre-stress levels rather than any standard levels, as plasma glucose is also dependent on feeding status, diet type, and other factors (26). However, overall glucose levels may convey important information in terms of severe stress. According Martínez-Porchas et al. (60), glucose cannot be eliminated from a list of stress indicators, preferably when evaluating a chronic exposure to stressful conditions, but must be complemented with other stress measurements as hormones or blood-cell counts, in order to have a more complete profile about the stress status of any fish.

High rusticity is associated with tilapia, involving ample tolerance to a wide range of temperature and pH. Thus, predictably, most results for environmental indicators were within acceptable levels for the species. However, DO on farms A and B was critically low. Despite the species strong ability to survive a few hours even under anoxia, DO is one of the limiting environmental factors for tilapia (61). The concentration of DO in water is influenced by water temperature, salinity and atmospheric pressure (62), and the solubility of oxygen decreases as temperature increases (63). Ross and Ross (64) reported that tilapia handling increased DO consumption from 150 to 300% and that the high temperature also increases the oxygen consumption. All these factors may have contributed to the extremely DO low rates observed in farms A and B. Therefore, specific stress factors must be avoided in warm periods. Also, in the three scenarios, it was difficult to establish water renewal rates, as there was practically no inflow, due to the abnormally dry summer that occurred in the South and Southeast regions of Brazil. The chronic exposure to low DO causes fish immunosuppression and performance reduction (51), affecting welfare direct and indirectly. When DO reaches 45 to 50% of saturation (3.0–3.5 mg/L, at 28–30°C), tilapia reduces it metabolic activity as a regulation mechanism, reducing respiration and growth, and in saturations between 10 and 20% (0.7–1.6 mg/L at 26–35°C) generate great discomfort and eventual mortality (34, 65). The chronic stress associated with the metabolic reduction caused by the low oxygen availability may affect fish growth (66), which may explain the negative correlation we observed between DO and K factor. Normally, in Brazil the tilapia production cycle occurs in 6–9 months, with a target body weight of around 800 g (34, 67); however, in farm A, the animals were older than 12 months. The very low DO and chronic stress related to this indicator may have interfered in farmed tilapia growing ratio. Slow growth was also observed on farm C, probably due to malnutrition caused by inadequate food distribution, aligned with the observation of fast feed consumption in most cages. On farm B, slow growth rate was not observed, despite a low oxygen level, a result likely related to the young age of the animals and, consequently, no possibility for the evaluation of growth rate within a longer period of time. In addition, on farm B fish were apparently well-nourished, based on fish weight and K factor, and the feed was provided in adequate quantities. The K factor was an important indicator of nutritional status with discriminating power between different growth stages of farms, as it is inversely proportional to fish length, explaining the significance between farms A and B. Standard K factor for tilapia was not found in the literature; however, our results seem coherent with previously reported values. Ighwela et al. (68) reported K factor varying between 1.64 and 1.79 for tilapia fingerlings of 14.52 ± 6.39 g fed on different maltose levels; Anani and Nunoo (69) founded a K factor of 2.01 for fish weighing 140.3 ± 23 g when consuming a specific formulated diet. These results also show variation in K factor according to fish development stage, a characteristic that requires attention.

The economic demand for a short production cycle and fast growth rate was likely related to the high protein levels in the fish diet. However, Mengistu et al. (40) showed that tilapia FCR decreased with increasing CP, DO, and pH. Thus, this management may negatively affect fish FCR and their weight gain. Excessive protein levels result in additional energy expenditures, as excess amino acids require metabolization (70). The integrated production system adopted in the State of Santa Catarina, in which the slaughterhouse supplies the fingerlings and the feed to producers, may influence feeding decisions, as producers are exempt from feeding costs, but committed to delivering fish for slaughter in a short period of time. Decisions regarding feed distribution management are also extremely important (1). When feed is offered exclusively from one pond margin side, a privilege based on behavioral dominance is favored, with the larger animals becoming better fed and the smaller animals prevented from accessing adequate amounts of feed (71). In cages, fish that are located in the superficial water column tend to be benefited (71). This is one of the reasons for unevenness in fish weight and the occurrence of many rejected animals on farm C. When animals cannot satisfy their motivation for feeding, their welfare is compromised. As a source of additional secondary welfare problems, underfeeding tends to increase agonistic behavior due to the intense competition for resource, which may result in injuries (19). The use of demand-feeders spread on the pond margins reaching different depths in the case of cages, may be an alternative for underfeeding (51). Nevertheless, to ensure adequate nutrition, it is necessary to study the distribution of tilapias in the water column and the fish dominance behavior during feeding.

During the catabolism of ingested proteins, fish produce nitrogenous waste which is excreted through urine. The main end product of such catabolic activity is ammonia, which is toxic for fish. Ammonia is also derived from decomposition of organic material such as feed leftovers, feces, and organic fertilizer (72). However, its toxicity depends on other water physicochemical parameters, mainly pH and temperature (73). As pH increases above 7.0, a greater percentage of total ammonia is converted from the ionic form (NH_4) to the toxic un-ionized gaseous form (NH_3) (74). In addition, ammonia is more toxic at higher temperatures (75). Despite the accumulation of feed in some locations on farm C and the low water renewal, high values of NH_3 and NO_2 were not observed. As on farm A and B pH was acid and neutral, respectively, the ammonia value registered did not represent risk of toxicity.

Low alkalinity observed on farm A is associated with water acidification, due to the lack of buffering capacity of the system. When bicarbonates and carbonates are maintained in satisfactory levels, the pH tends to be stable, avoiding fish acidic or alkaline stress (51). The total alkalinity of water tends to be higher with the presence of phytoplankton (green waters), due to the consumption of CO_2 by the algae (76). The phytoplankton is indirectly measured by the water transparence, and when maintained in equilibrium is an important additional food source for tilapia (77, 78). Besides that, algae may minimize the incidence of excessive light incidence in ponds (79). Thus, the maintenance of adequate levels of phytoplankton improves water quality and may provide greater comfort to fish in relation to environment luminosity, potentially improving tilapia welfare.

The external environment also influences fish welfare. For example, predation can be a cause of high mortality and stress in farmed fish. According to Broom (1), when exposed to predators, fish can show strong emergency adrenal responses and suppression of feeding behavior. Prevention measures for predator control include netting above or inside water, acoustic or visual devices (80, 81). On farm C, cages were screen-covered; however, birds were frequently present and often landed on that, suggesting that further actions are required. Anti-predator strategies are needed in excavated ponds, and because no actions were observed to minimize predation, this seems a relevant critical point for welfare assessment. An additional external deleterious effect on welfare was the excess of light exposure, since none of ponds had any shielding from direct sun light. A possible solution is the use of a fine mash above ponds or cages. The shading promoted by cage-covers can minimize the ultraviolet light and fright stress caused by overhead movement, both of which tend to reduce the risk of chronic stress and avoid predatory birds (74). Excessive lighting that occurs during massive capture operations is also a stressful factor, and when prolonged it becomes proportionally more deleterious to fish welfare.

Despite being a potential stress factor, air exposure, which occurred during pre-slaughter management, was trivialized at all facilities visited. The fact that tilapia can survive out of water for some period does not mean that it is a stress-free experience (47). The damages caused by air exposure depend on its duration and the fish species. European Food Safety Authority (EFSA) recommendation for trout is that air exposure be at most 10 s (82). This duration allowance for air exposure is critically inferior to the duration of at least 10 min observed in tilapia capture and pre-slaughter procedures on the studied farms. This confirms the urgent necessity of establishing welfare guidelines for tilapia farming.

In general, massive capture is associated with crowding, air and light exposure. The fish physiological response to these acute handling stressors is altered by previous long-term holding conditions (83). Crowding procedures are improved by efforts of reducing their duration and severity, in order to avoid additional suffering, stress, injuries, and mortality (5, 9, 84, 85). Considering the extremely high level and duration of crowding adopted during massive capture on farm A, that handling qualifies as unacceptable in terms of fish welfare, according to classification proposed by Noble et al. (26). Currently, there are management alternatives that avoid the contact of fish with air and light, through mechanical pumping (45). However, the main operational difficulty of mechanical pumping is to assure the removal of the whole population, including fish at the bottom of the tank and the injuries caused by suction pressure (47). Considering the traditional method, the color, size, and material of the sweep net may influence the stress caused to fish during capture, and the best choice of equipment depends on the situation (5). Lines and Spence (47) stated that welfare at capture can be improved by adopting and adapting procedures used for other species or developing completely new concepts or methods. Therefore, to mitigate stressors during tilapia capture, it is preferred to adopt the procedures and equipment that result in faster capture with less abrasive material, causing the lowest level of crowding, depending on circumstances.

Slaughter is considered one of the main critical points for fish welfare, mainly due to the lack of standardization and of legislation on fish humane slaughter practices (86, 87). The three main indicators of humane slaughter are the avoidance of excitement, pain, and suffering in the pre-slaughter handling, the loss of pain sensitivity within <1 s of the application of any aversive stunning or slaughter procedure, and its persistence until death (16). Methods of asphyxiation, decapitation with adherence of organs, and the pre-slaughter or slaughter with ice slurry observed in this work do not promote instant unconsciousness. Therefore, these methods cannot be considered humane (5, 9, 15). Including, there was resistance to allow the monitoring of these practices in at least six properties visited during the construction of this protocol. This may indicate that people are insecure about the adequacy of the practices adopted. According to Pedrazzani et al. (20), 87% of people interviewed in the town of Araucária, Southern Brazil, believed that fish are capable to feel pain and 85% that common slaughter methods cause suffering. Similar results were obtained by Rucinque et al. (88), who conducted an interview with highly educated citizens from Bogotá and Curitiba. From the participants, 79.7 and 71.8% perceived fish as sentient animals, and 76.0 and 72.0% believed that fish should be included in humane slaughter regulations, respectively. Webster (89) suggested that there is a gradual acceptance by farmers, scientists and veterinarians that farmed fish need to be treated in a humane and compassionate manner. As for slaughter, some efforts have been applied in the development of humane methods using electrical stunning for tilapia in Brazil; however, there is uncertainty regarding its effectiveness, due the lack of the monitoring of fish consciousness specific technical support and data registers during process.

Finally, general questions regarding overall welfare management are relevant. Farms in low-standard conditions are generally at greater risk of failing to respond to the basic welfare needs of farmed fish (90). This may be worse if there is a prevailing understanding that the animals are biologically able to cope with captive conditions, as is the common perception regarding tilapia. Even sturdier species may suffer with environmental challenges and it is the attribution of those responsible for the animals to always seek the best maintenance and management conditions. In this sense, tilapia welfare assessment may be used to identify critical welfare areas to be improved on farm (32). Some tilapia critical welfare points were common across all farms, even though the visited farms were diverse, including in terms of production systems adopted. These transversal critical welfare issues were the low rates of DO in water, the long duration of management for fish capture with exposure to air and crowding, and strongly aversive slaughter methods, which cannot be considered humane.

The lack of statistical significance between farms when comparing total scores is probably associated with the need for improvement in the integration of individual scores into a final overall welfare category to each farm. This is a recurrent difficulty in animal welfare assessment, to which even more refined integration methods, such as those proposed by the

Welfare Quality protocols (22, 23, 91), have not yet provided a completely satisfactory solution (92). Even though the main goal of our work was to determine a robust protocol containing major critical welfare points in different scenarios, we believe further research into the integration of individual scores to produce an overall welfare assessment warrants further studies.

APPLICATIONS AND CONCLUSIONS

The identification of critical tilapia welfare issues seems essential for farmers to adopt preventive management actions (93). For example, some conditions such as gill and fin problems are affected not just by handling, but also by the confinement conditions (5, 44) and seem to deserve higher levels of attention. Then, the regular use of a tilapia welfare assessment protocol becomes an important management tool. Additionally, the protocol is open to the inclusion of new welfare indicators, and the enrichment of the list of behavioral indicators is urgent, especially indicators that allow for a closer observation of tilapia behavior throughout their complete life cycles. This priority is evident, as the understanding of welfare depends inherently on the direct observation of the individuals and animal behavior is a major form for the expression of emotions and feelings in non-verbal species (94). Despite the challenge of high turbidity of pond water, methods for underwater behavioral assessment must be developed in order to obtain a better understanding of specific issues such as the hierarchical relationship between fish, especially during the feeding; the occupation of the water column in terms of cage or pond area that is actually useful for the tilapias and its implications for natural swimming behavior; the proper calculation of density; and the development of environmental enrichment techniques (90). In few words, in order to be effective in the monitoring and enhancement of animal welfare, animals must be seen throughout their lives.

It seems relevant to emphasize that the practical application of this first protocol, even though it is not exhaustive, will allow the producer to be closer to the animals, just as it happens with terrestrial vertebrates. This strengthening of communication may be an ally to the prevention of diseases and control of other potential problems related to water quality, external environment, and inadequate management, thus minimizing the harmful effects caused by the productive systems to the welfare of tilapia. This may be an initial step for a tilapia welfare strategy, where the prioritization of critical points, implementation of corrective actions and monitoring of the results is part of a permanent welfare management program. A final important remark is the fact that the protocol also lends itself to adaptation into a mobile application, which may further facilitate on-farm use and promote its adoption.

Our results suggest that a tilapia welfare assessment routine may be in place with a single protocol, which seems effective in different farming realities and feasible for farm staff use. Furthermore, the developed protocol has shown relative discriminating power, high on-field feasibility and a clear role in determining critical points in tilapia welfare, which in turn may guide management decisions. Considering the challenges presented for further improvements to the protocol, we believe that the format presented, which is compatible with and close to that of other species welfare assessment protocols with longer history of use and refinements may help the identification of best future approaches. Finally, refinements to the protocol are welcome in relation to the integration of the indicators into a single final score for each property, in addition to the continuous refinement of the existent indicators and the inclusion of new tilapia welfare indicators as they become available.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The animal study was reviewed and approved by Animal Use Ethics Committee of the Agricultural Campus, Federal University of Paraná (No. 083/2019).

AUTHOR CONTRIBUTIONS

MQ: funding acquisition. AP, CM, and MQ: study conception and design. AP, FB, and ES: performance of the data collection and tabulation. AP: statistical modeling and data analyses. AP and CM: preparation of the manuscript. MQ and CM: project coordination. All authors reviewed and approved the final version of manuscript.

FUNDING

This work completely funding was provided by FAI Farms Limited, including one postdoctoral fellowship in Veterinary Science and one Veterinary Medicine scientific initiation undergraduate grant.

ACKNOWLEDGMENTS

We thank Fernanda Queiróz and Silva and Susane Pahl-Klipp from the Rural Development Unit Secretariat of Agriculture and Environment (SAMA) of Joinville, Tatiane Carine da Silva, Bruno Corrêa da Silva, and Hilton Amaral Junior from Santa Catarina Agricultural Research and Rural Extension Company (Epagri) for providing access to tilapia farms and field support, and all farmers involved for their trust and permission to conduct this job, Maria Eduarda Gurski Coelho, undergraduate student of the Veterinary Medicine, for her participation on field data collection. Finally, we would like to acknowledge the suggestions provided by the two reviewers, which significantly contributed to improve of our manuscript.

REFERENCES

1. Broom DM. Cognitive ability and sentience: which aquatic animals should be protected? *Dis Aquat Organ.* (2007) 75:99–108. doi: 10.3354/dao075099
2. Chandroo KP, Duncan IJH, Moccia RD. Can fish suffer?: Perspectives on sentience, pain, fear and stress. *Appl Anim Behav Sci.* (2004) 86:225–50. doi: 10.1016/j.applanim.2004.02.004
3. Braithwaite V. *Do Fish Feel Pain?* Oxford: Oxford University Press (2010). p. 208.
4. Sneddon LY, Lopez-Luna J, Wolfenden DCC, Leach MC, Valentim AM, Steenbergen PJ, et al. Fish sentience denial: muddying the waters. *Anim Sentience.* (2018) 115:1–12.
5. Branson EJ. *Fish Welfare*. Oxford: Blackwell Publishing (2008). p. 312.
6. Broom DM, Johnson KG. *Stress and Animal Welfare*. Dordrecht: Springer (1993). p. 228.
7. Conte FS. Stress and the welfare of cultured fish. *Appl Anim Behav Sci.* (2004) 86:205–23. doi: 10.1016/j.applanim.2004.02.003
8. Galhardo L, Oliveira R. Bem-estar animal: um conceito legítimo para peixes? *Rev Etol.* (2006) 8:51–61.
9. Håstein T, Scarfe AD, Lund VL. Science-based assessment of welfare: aquatic animals. *Rev Sci Tech.* (2005) 24:529–47. doi: 10.20506/rst.24.2.1590
10. Lymbery P. *In Too Deep the Welfare of Intensively Farmed Fish a Report for Compassion in World Farming Trust*. Petersfield: Compassion in World Farming Trust (2002). p. 65.
11. Schwedeler TE, Johnson SK. Responsible care and health maintenance of fish in commercial aquaculture. *Anim Welf Information Centre Bull.* (2000) 10:3–4.
12. Bhujel RC. A review of strategies for the management of Nile tilapia (*Oreochromis niloticus*) broodfish in seed production systems, especially hapa-based systems. *Aquaculture.* (2000) 181:37–59. doi: 10.1016/S0044-8486(99)00217-3
13. Barcellos LJG, Nicolaiewsky S, De Souza SMG, Lulhier F. Plasmatic levels of cortisol in the response to acute stress in Nile tilapia, *Oreochromis niloticus* (L.), previously exposed to chronic stress. *Aquac Res.* (1999) 30:437–44. doi: 10.1046/j.1365-2109.1999.00348.x
14. FSBI. *Fish Welfare. Briefing Paper 2, Fisheries Society of the British Isles Granta Information Systems*. Cambridge: Granta Information System (2002). p. 21.
15. Robb D, Kestin SC. Methods used to kill fish: field observations and literature reviewed. *Anim Welf.* (2002) 11:269–82.
16. Lines JA, Robb DH, Kestin SC, Crook SC, Benson T. Electric stunning: a humane slaughter method for trout. *Aquac Eng.* (2003) 28:141–54. doi: 10.1016/S0144-8609(03)00021-9
17. Van de Vis H, Kestin S, Robb D, Oehlenschläger J, Lambooij B, Münkner W, et al. Is humane slaughter of fish possible for industry? *Aquac Res.* (2003) 34:211–20. doi: 10.1046/j.1365-2109.2003.00804.x
18. Boyland N. *The Welfare of Farmed Fish During Slaughter in the European Union*. Compassion in World Farming International (2018). p. 52.
19. Ellis T, Oidtmann B, St-Hilaire S, Turnbull J, North B, Mac- Intyre C, et al. Fin erosion in farmed fish. In: Branson EJ, editor. *Fish Welfare*. Oxford: Blackwell (2008). p. 121–49.
20. Pedrazzani AS, Carneiro PCF, Kirschnik PG, Molento CFM. Impacto negativo de secção de medula e termonarcose no bem-estar e na qualidade da carne da tilápia-do-Nilo. *Rev Bras Saúde e Produção Anim.* (2009) 10:188–97.
21. FAWC (Farmed Animal Welfare Council). *Report on the Welfare of Farmed Fish*. Farm Animal Welfare Committee, London (1996). p. 43.
22. Blokhuis HJ, Veissier I, Miele M, Jones B. The welfare quality® project and beyond: safeguarding farm animal well-being. *Acta Agric Scand A Anim Sci.* (2010) 60:129–40. doi: 10.1080/09064702.2010.523480
23. Battini M, Stilwell G, Vieira A, Barbieri S, Canali E, Mattiello S. On-farm welfare assessment protocol for adult dairy goats in intensive production systems. *Animals.* (2015) 5:934–50. doi: 10.3390/ani5040393
24. Dwyer C, Ruiz R, Heredia IB de, Canali E, Barbieri S, Zanella A. *AWIN Welfare Assessment Protocol for Sheep*. (2015). p. 72. doi: 10.13130/AWIN_sheep_2015
25. Ferrante V, Watanabe TTN, Marchewka J, Estevez I. *AWIN Welfare Assessment Protocol for Turkeys*. (2015). p. 45. doi: 10.13130/AWIN_TURKEYS_2015
26. Noble C, Gismervik K, Iversen MH, Kolarevic J, Nilsson J, Stien LH, et al. *Welfare Indicators for Farmed Atlantic Salmon: Tools for Assessing Fish Welfare*. Tromsø: Nofima (2018). p. 351.
27. Scott E, Nolan A, Fitzpatrick J. Conceptual and methodological issues related to welfare assessment: a framework for measurement. *Acta Agric Scand Sect A Anim Sci Suppl.* (2001) 30:5–10. doi: 10.1080/090647001316922983
28. Whay H, Main D, Green L, Webster A. Animal-based measures for the assessment of welfare state of dairy cattle, pigs and laying hens: consensus of expert opinion. *Anim Welf.* (2003) 12:205–17.
29. Knierim U, Winckler C. On-farm welfare assessment in cattle: validity, reliability and feasibility issues and future perspectives with special regard to the Welfare Quality® approach. *Anim Welf.* (2009) 18:451–8.
30. Pettersen JM, Bracke MBM, Midtlyng PJ, Folkedal O, Stien LH, Steffenak H, et al. Salmon welfare index model 2.0: an extended model for overall welfare assessment of caged Atlantic salmon, based on a review of selected welfare indicators and intended for fish health professionals. *Rev Aquac.* (2014) 6:162–79. doi: 10.1111/raq.12039
31. Stien LH, Bracke MBM, Folkedal O, Nilsson J, Oppedal F, Torgersen T, et al. Salmon Welfare Index Model (SWIM 1.0): a semantic model for overall welfare assessment of caged Atlantic salmon: review of the selected welfare indicators and model presentation. *Rev Aquac.* (2013) 5:33–57. doi: 10.1111/j.1753-5131.2012.01083.x
32. RSPCA. *RSPCA Welfare Standards for Farmed Atlantic Salmon*. (2018). Available online at: https://www.berspcaassured.org.uk/media/1290/rspca-welfare-standards-for-salmon-february-2018.pdf (accessed March 20, 2019).
33. RSPCA. *RSPCA Welfare Standards for Farmed Rainbow Trout*. (2018). Available online at: https://science.rspca.org.uk/sciencegroup/farmanimals/standards/trout (accessed March 20, 2019).
34. Kubitza F, Kubitza LMM. Tilápias: qualidade da água, sistemas de cultivo, planejamento da produção, manejo nutricional e alimentar e sanidade – parte I. *Panor Aquicul.* (2000) 10:45–7.
35. Emerson K, Russo RC, Lund RE, Thurston RV. Aqueous ammonia equilibrium calculations: effect of pH and temperature. *J Fish Res Board Canada.* (1975) 32:2379–83. doi: 10.1139/f75-274
36. Davis MW. Fish stress and mortality can be predicted using reflex impairment. *Fish Fish.* (2010) 11:1–11. doi: 10.1111/j.1467-2979.2009.00331.x
37. Kestin SC, Van de Vis JW, Robb DHF. Protocol for assessing brain function in fish and the effectiveness of methods used to stun and kill them. *Vet Rec.* (2002) 150:302–7. doi: 10.1136/vr.150.10.302
38. Damsgård B, Juell J, Braastad BO. *Welfare in Farmed Fish. Fiskeriforskning*. Tromsø: Norwegian Institute of Fisheries and Aquaculture Research (2006). p. 105.
39. Souza AP de O, Molento CFM. Proposal of a management system to develop an animal welfare strategy for the animal food chain. *CAB Rev Perspect Agric Vet Sci Nutr Nat Resour.* (2018) 13:1–11. doi: 10.1079/PAVSNNR201813001
40. Mengistu SB, Mulder HA, Benzie JAH, Komen H. A systematic literature review of the major factors causing yield gap by affecting growth, feed conversion ratio and survival in Nile tilapia (*Oreochromis niloticus*). *Rev Aquac.* (2020) 12:524–41. doi: 10.1111/raq.12331
41. Smith SA. *Fish Diseases and Medicine*. Boca Raton, FL: CRC Press (2019). p. 414.
42. Gregory NG, Grandin T. *Animal Welfare and Meat Production*. Oxon: CABI Publishing (2007). p. 299.
43. Becerra J, Montes GS, Bexiga SRR, Junqueira LCU. Structure of the tail fin in teleosts. *Cell Tissue Res.* (1983) 230:127–37. doi: 10.1007/BF00216033
44. Noble C, Jones HAC, Damsgård B, Flood MJ, Midling KO, Roque A, et al. Injuries and deformities in fish: their potential impacts upon aquacultural production and welfare. *Fish Physiol Biochem.* (2012) 38:61–83. doi: 10.1007/s10695-011-9557-1
45. Southgate P, Wall T. Welfare of farmed fish at slaughter. *In Pract.* (2001) 23:277–84. doi: 10.1136/inpract.23.5.277

46. Barthel BL, Cooke SJ, Suski CD, Philipp DP. Effects of landing net mesh type on injury and mortality in a freshwater recreational fishery. *Fish Res.* (2003) 63:275–82. doi: 10.1016/S0165-7836(03)00059-6
47. Lines JA, Spence J. Safeguarding the welfare of farmed fish at harvest. *Fish Physiol Biochem.* (2012) 38:153–62. doi: 10.1007/s10695-011-9561-5
48. Jurk I. Ophthalmic disease of fish. *Vet Clin North Am Exot Anim Pract.* (2002) 5:243–60. doi: 10.1016/S1094-9194(02)00006-3
49. Zamri-Saad M, Amal MNA, Siti-Zahrah A. Pathological changes in red tilapias (*Oreochromis* spp.) naturally infected by Streptococcus agalactiae. *J Comp Pathol.* (2010) 143:227–9. doi: 10.1016/j.jcpa.2010.01.020
50. Jansen MD, Mohan CV. *Tilapia lake Virus (TiLV): Literature Review*. Bayan Lepas: CGIAR Research Program on Fish Agri-Food Systems (2017). p. 12.
51. El-Sayed A-FM. *Tilapia Culture.* Wallingford, CT: CABI Publishing (2006). p. 277.
52. Magnoni LJ, Martos-Sitcha JA, Prunet P, Mancera JM. Editorial: welfare and stressors in fish: challenges facing aquaculture. *Front Physiol.* (2020) 11:162. doi: 10.3389/fphys.2020.00162
53. Mori RH, Chedid RA, Braccini GL, Ribeiro RP, Lopes De Oliveira CA, Pretto-Giordano LG, et al. Prevalence of ectoparasites and bacteriological diagnosis in Nile tilapia bred in net-tanks in the Corvo's river, Paraná, Brazil. *Semin Agrar.* (2015) 36:1145–54. doi: 10.5433/1679-0359.2015v36n2p1145
54. Ranzani-Paiva MJT, Nunes Felizardo N, Luque JL. Parasitological and hematological analysis of Nile tilapia *Oreochromis niloticus* Linnaeus, 1757 from Guarapiranga reservoir, São Paulo State, Brazil. *Acta Sci Biol Sci.* (2005) 27:231–7. doi: 10.4025/actascibiolsci.v27i3.1334
55. Vargas L, Povh JA, Ribeiro RP, Moreira HLM. Ocorrência de ectoparasitos em tilápias do Nilo (*Oreochromis niloticus*), de origem tailandesa, em Maringá-Paraná. *Arq Ciên Vet Zool Unipar.* (2000) 3:31–7.
56. Zago AC, Franceschini L, Garcia F, Schalch SHC, Gozi KS, Silva RJ da. Ectoparasites of Nile tilapia (*Oreochromis niloticus)* in cage farming in a hydroelectric reservoir in Brazil. *Rev Bras Parasitol Vet.* (2014) 23:171–8. doi: 10.1590/S1984-29612014041
57. De Jong IC, Hindle VA, Butterworth A, Engel B, Ferrari P, Gunnink H, et al. Simplifying the Welfare Quality® assessment protocol for broiler chicken welfare. *Animal.* (2016) 10:117–27. doi: 10.1017/S1751731115001706
58. Bui S, Oppedal F, Sievers M, Dempster T. Behaviour in the toolbox to outsmart parasites and improve fish welfare in aquaculture. *Rev Aquac.* (2019) 11:168–86. doi: 10.1111/raq.12232
59. Barber I. Parasites, behaviour and welfare in fish. *Appl Anim Behav Sci.* (2007) 104:251–64. doi: 10.1016/j.applanim.2006.09.005
60. Martínez-Porchas M, Martínez-Córdova LR, Ramos-Enriquez R. Cortisol and glucose: reliable indicators of fish stress? *Pan Am J Aquat Sci.* (2009) 4:158–78.
61. Ross LG. Environmental physiology and energetics. In: Beveridge MCM, McAndrew B, editors. *Tilapias: Biology and Exploitation.* Norwell, MA: Kluwer Academic Publishers (2000). p. 89–128.
62. Wetzel RG. *Limnology: Lake and River Ecosystems*. San Diego, CA: Academic Press (2001). p. 1006.
63. Eaton DE, Clesceri LS, Rice EW, Greenberg AE, Franson MAH. *Standard Methods for the Examination of Water & Wastewater*. Washington, DC: American Public Health Association (2005). p. 1220.
64. Ross B, Ross LG. The oxygen requirements of *Oreochromis niloticus* under adverse conditions. In: Fishelson L, Yaron Z, editors. *Proceedings of the International Symposium on Tilapia in Aquaculture*, 8–13. May 1983. Nazareth: Tel Aviv University, Israel (1983). p. 134–43.
65. Kubitza LMM, Guimarães TG, Kubitza F. Monitorando a Saúde Dos Peixes. *Panor da Aquicultura* (2000) 10:32—9.
66. Nehemia A, Maganira J. Length-Weight relationship and condition factor of tilapia species grown in marine and fresh water ponds. *Agric Biol J North Am.* (2012) 3:117–24. doi: 10.5251/abjna.2012.3.3.117.124
67. Silva BC, Massago H, Marchiori NC. *Monocultivo de Tilápia em Viveiros Escavados em Santa Catarina*. Itajaí: Epagri (2019). p. 128.
68. Ighwela KA, Ahmed AB, Abol-Munafi AB. Condition factor as an indicator of growth and feeding intensity of Nile Tilapia fingerlings (*Oreochromis niloticus*) feed on different levels of maltose. *Am J Agric Environ Sci.* (2011) 11:559–63.
69. Anani FA, Nunoo FKE. Length-weight relationship and condition factor of Nile tilapia, *Oreochromis niloticus* fed farm-made and commercial tilapia diet. *Int J Fish Aquat Stud.* (2016) 4:647–50.
70. Carneiro WF, Pandini F, da Silva LCR, Dos Santos LD, Rossato KA, Meurer F. Exigência de proteína digestível para tilápias do Nilo alimentadas com rações à base de farelo de soja e milho. *Acta Sci Anim Sci.* (2017) 39:343–9. doi: 10.4025/actascianimsci.v39i4.36122
71. Attia J, Millot S, Di-Poï C, Bégout ML, Noble C, Sanchez-Vazquez FJ, et al. Demand feeding and welfare in farmed fish. *Fish Physiol Biochem.* (2012) 38:107–18. doi: 10.1007/s10695-011-9538-4
72. Wicks BJ, Randall DJ. The effect of feeding and fasting on ammonia toxicity in juvenile rainbow trout, Oncorhynchus mykiss. *Aquat Toxicol.* (2002) 59:71–82. doi: 10.1016/S0166-445X(01)00237-5
73. Bower CE, Bidwell JP. Ionization of ammonia in seawater: effects of temperature, pH, and salinity. *J Fish Res Board Canada.* (1978) 35:1012–6. doi: 10.1139/f78-165
74. Webster CD, Lim C. *Tilapia: Biology, Culture, and Nutrition.* Binghamton: Food Products Press (2006). p. 704.
75. Boyd CE, Tucker CS. *Pond Aquaculture Water Quality Management.* Norwell, MA: Kluwer Academic Publishers (1998). p. 700.
76. Cavalcante DH, Do Carmo e Sá MV. Efeito da fotossíntese na alcalinidade da água de cultivo da tilápia do Nilo. *Rev Cienc Agron.* (2010) 41:67–72.
77. Athanasio SM, Yunus DM, Samwel ML, Amon PS. Phytoplankton and food selectivity in Nile tilapia reared in earthen ponds under monoculture and polyculture with African Sharptooth catfish. *J Aqua Trop.* (2017) 32:15–38.
78. Moriarty CM, Moriarty DJW. Quantitative estimation of the daily ingestion of phytoplankton by *Tilapia nilotica* and *Haplochromis nigripinnis* in Lake George, Uganda. *J Zool.* (1973) 171:15–23. doi: 10.1111/j.1469-7998.1973.tb07513.x
79. Leira MH, Cunha LT da, Braz MS, Melo CCV, Botelho HA, Reghim LS. Qualidade da água e seu uso em pisciculturas. *Pubvet.* (2017) 11:11–7. doi: 10.22256/pubvet.v11n1.11-17
80. Amos N, Sullivan R. *The Business of Farm Animal Welfare.* Routledge: Greenleaf Publishing (2017). p. 304.
81. Guerrero RD. *Coping Strategies for Climate Change Impacts on Philippine Aquaculture.* NAST Bulletin n°. 11. Bicutan: National Academy of Science and Technology (2017). p. 4.
82. Algers B, Blokhuis HJ, Bøtner A, Broom DM, Costa P, Domingo M, et al. Scientific opinion. Species-specific welfare aspects of the main systems of stunning and killing of farmed carp Scientific Opinion of the Panel on Animal Health and Welfare (Question N° EFSA-Q-2008-439). *EFSA J.* (2009) 1013:1–37.
83. Barton BA, Ribas L, Acerete L, Tort L. Effects of chronic confinement on physiological responses of juvenile gilthead sea bream, *Sparus aurata* L., to acute handling. *Aquac Res.* (2005) 36:172–9. doi: 10.1111/j.1365-2109.2004.01202.x
84. Bagni M, Civitareale C, Priori A, Ballerini A, Finoia M, Brambilla G, et al. Pre-slaughter crowding stress and killing procedures affecting quality and welfare in sea bass (*Dicentrarchus labrax)* and sea bream (*Sparus aurata*). *Aquaculture.* (2007) 263:52–60. doi: 10.1016/j.aquaculture.2006.07.049
85. Skjervold PO, Fjæra SO, Østby PB, Olai E. Live-chilling and crowding stress before slaughter of Atlantic salmon Salmo salar. *Aquaculture.* (2001) 192:265–80. doi: 10.1016/S0044-8486(00)00447-6
86. Viegas EMM, Pimenta FA, Previero TC, Gonçalves LU, Durães JP, Ribeiro MAR, et al. Métodos de abate e qualidade da carne de peixe. *Arch Zootec.* (2011) 61:41–50. doi: 10.21071/az.v61i237.2957
87. Ferreira N de A, Araújo RV de, Campos EC. Boas práticas no pré-abate e abate de pescado. *Pubvet.* (2018) 12:1–14. doi: 10.31533/pubvet.v12n7a137.1-14
88. Rucinque DS, Souza APO, Molento CFM. Perception of fish sentience, welfare and humane slaughter by highly educated citizens of Bogotá, Colombia and Curitiba, Brazil. *PLoS ONE.* (2017) 12:e0168197. doi: 10.1371/journal.pone.0168197
89. Webster J. *Management and Welfare of Farm Animals: The UFAW Farm Handbook.* Oxford: Wiley-Blackwell (2011). p. 614.
90. Martins CIM, Galhardo L, Noble C, Damsgård B, Spedicato MT, Zupa W, et al. Behavioural indicators of welfare in farmed fish. *Fish Physiol Biochem.* (2012) 38:17–41. doi: 10.1007/s10695-011-9518-8

91. Dalmau A, Temple D, Rodríguez P, Llonch P, Velarde A. Application of the Welfare Quality® protocol at pig slaughterhouses. *Anim Welf.* (2009) 18:497–505.
92. Buijs S, Ampe B, Tuyttens FAM. Sensitivity of the Welfare Quality® broiler chicken protocol to differences between intensively reared indoor flocks: which factors explain overall classification? *Animal.* (2017) 11:244–53. doi: 10.1017/S1751731116001476
93. Saraiva JL, Arechavala-Lopez P, Castanheira MF, Volstorf J, Studer BH. A global assessment of welfare in farmed fishes: the fishethobase. *Fishes.* (2019) 4:1–18. doi: 10.3390/fishes4020030
94. Anderson DJ, Adolphs R. A framework for studying emotions across species. *Cell.* (2014) 157:187–200. doi: 10.1016/j.cell.2014.03.003

Conflict of Interest: MQ and FB were employed by FAI Farms.

The remaining 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.

frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 30 November 2020
doi: 10.3389/fvets.2020.580523

The Influence of Predictability and Controllability on Stress Responses to the Aversive Component of a Virtual Fence

Tellisa Kearton[1,2], Danila Marini[1,2], Frances Cowley[1], Sue Belson[2], Hamideh Keshavarzi[2], Bonnie Mayes[1,2] and Caroline Lee[1,2*]

[1] School of Environmental and Rural Science, University of New England, Armidale, NSW, Australia, [2] Agriculture and Food, Commonwealth Scientific and Industrial Research Organization, Armidale, NSW, Australia

To ensure animal welfare is not compromised, virtual fencing must be predictable and controllable, and this is achieved through associative learning. To assess the influence of predictability and controllability on physiological and behavioral responses to the aversive component of a virtual fence, two methods of training animals were compared. In the first method, positive punishment training involved sheep learning that after an audio stimulus, an electrical stimulus would follow only when they did not respond by stopping or turning at the virtual fence (predictable controllability). In the second method, classical conditioning was used to associate an audio stimulus with an electrical stimulus on all occasions (predictable uncontrollability). Eighty Merino ewes received one of the following treatments: control (no training and no stimuli in testing); positive punishment training with an audio stimulus in testing (PP); classical conditioning training with only an audio stimulus in testing (CC1); and classical conditioning training with an audio stimulus followed by electrical stimulus in testing (CC2). The stimuli were applied manually with an electronic collar. Training occurred on 4 consecutive days with one session per sheep per day. Sheep were then assessed for stress responses to the cues by measuring plasma cortisol, body temperature and behaviors. Predictable controllability (PP) sheep showed no differences in behavioral and physiological responses compared with the control treatment ($P < 0.05$). Predictable uncontrollability of receiving the aversive stimulus (CC2) induced a higher cortisol and body temperature response compared to the control but was not different to CC1 and PP treatments. CC2 treatment sheep showed a higher number of turning behaviors ($P < 0.001$), and more time spent running ($P < 0.001$) than the control and PP treatment groups, indicating that predictability without controllability was stressful. The behavior results also indicate that predicting the event without receiving it (CC1) was less stressful than predicting the event then receiving it (CC2), suggesting that there is a cost to confirmation of uncontrollability. These results demonstrate that a situation of predictability and controllability such as experienced when an animal successfully learns to avoid the aversive component of a virtual fence, induces a comparatively minimal stress response and does not compromise animal welfare.

Keywords: animal welfare, behavior, cortisol, body temperature, electric shock, Bayesian brain, positive punishment

OPEN ACCESS

Edited by:
Edward Narayan,
The University of Queensland, Australia

Reviewed by:
Dominique Blache,
University of Western Australia, Australia
Manja Zupan Šemrov,
University of Ljubljana, Slovenia

****Correspondence:***
Caroline Lee
Caroline.lee@csiro.au

Specialty section:
This article was submitted to Animal Behavior and Welfare, a section of the journal Frontiers in Veterinary Science

Received: *06 July 2020*
Accepted: *30 October 2020*
Published: *30 November 2020*

Citation:
Kearton T, Marini D, Cowley F, Belson S, Keshavarzi H, Mayes B and Lee C (2020) The Influence of Predictability and Controllability on Stress Responses to the Aversive Component of a Virtual Fence. Front. Vet. Sci. 7:580523. doi: 10.3389/fvets.2020.580523

INTRODUCTION

The experience of stress in animals has psychological foundations, in which cognitive evaluation of the experience influences how stressful it is for the animal. In a series of experiments conducted in the 1970's, Weiss (1) demonstrated that the predictability and controllability of an electric shock influenced the degree of the stress response observed. Research has continued to investigate this phenomenon, with Greiveldinger et al. (2) finding that the predictability of a sudden event (sudden appearance of a panel above the feeding trough) reduced the number of startle responses observed and lambs showed less tachycardia when a light signal preceded the sudden event. The role of controllability of a stressor on animal welfare has been described in early work with rats using degree of gastric ulceration responses to electrical shocks (1, 3) which were reduced when the animals had predictability and controllability over their experience of the aversive event. Lambs which were taught to control an aversive event showed ear position and heart rate differences compared with control lambs, suggesting the perception of the aversive event was less stressful for lambs which could interrupt it (4). Long-term lack of predictability and controllability over stressors has been shown to lead to increased fearfulness in lambs as indicated through behavioral and physiological responses (5), as well as a pessimistic judgement bias (6). Predictability of stimuli have also been reviewed in relation to implications for captive animal welfare (7). The application of this research in a practical context has been investigated by Lee et al. (8) with the development of a framework in which the predictability and controllability of an animal's situation can be used in the assessment of the welfare state of an animal. The framework is based on the link between stress and welfare with the animals' cognitive evaluation of the predictability and controllability of the environment and their affective state resulting in positive or negative welfare outcomes. Stress responses result when animals are unable to predict or control negative events.

In the context of virtual fencing, associative learning is the mechanism through which an animal learns to avoid an aversive stimulus (an electrical stimulus applied through a collar) by responding to an audio stimulus (beep tone from the collar). This method is referred to as "positive punishment." In correctly responding to a benign audio cue (9) by either stopping forward movement or turning around, the animal successfully learns to avoid the aversive stimulus (10–13). Successful learning, therefore, implies that the animal learns to predict the occurrence of the aversive stimulus, and can control whether or not they receive the stimulus through their behavior. When an animal first encounters the virtual fence, the interaction is both unpredictable and uncontrollable and therefore has the potential to negatively impact welfare, so it is important to ensure that negative stimuli aren't so aversive as to create fear and distress. Positive punishment as a training technique has been utilized in numerous ways, commonly applied in horse training, in which aversive stimuli such as pressure from a whip, bit or spurs, encourages the animal to change its behavior in order to avoid receiving the aversive stimulus (14). The use of positive punishment has been criticized in dog (15), and horse training (16), due to complications with other training methods and inconsistencies in application of cues.

The perceptions of sheep to virtual fencing stimuli have been assessed in isolation with no prior experience in a previous study and it was found that the electrical stimulus was no more aversive than a commonly used restraint procedure with the audio cue being perceived as largely benign (9). To further test the welfare impacts of virtual fencing, the next step is to investigate the impact of these stimuli in relation to predictability and controllability. Successful learning of the virtual fencing system is proposed to be a predictable controllable situation, thereby inducing a minimal stress response to the audio cue following learning and reducing animal welfare risks (8). If the animal cannot predict or control receiving an aversive stimulus then its welfare is likely to be negatively impacted through increased fearfulness (5) and behavioral and physiological stress responses (6). Further, if the situation is on-going, negative states such as helplessness and hopelessness may result (1), with serious implications for animal welfare.

The first hypothesis of the study was that a capacity to predict and control the aversive (positive punishment) would eliminate the behavioral and physiological responses to the virtual fence and would not differ from the Control treatment. The second hypothesis was that a capacity to predict but not control the aversive stimulus (Classic Conditioning treatments) would induce a stress response and this would be greater in those animals receiving the aversive stimulus than those receiving the audio cue alone.

MATERIALS AND METHODS

The experiment was undertaken at CSIRO's McMaster Laboratory, Armidale, New South Wales (NSW), Australia. The protocol and conduct of the experiment was approved by the CSIRO Chiswick Animal Ethics Committee under the NSW Animal Research Act, 1985 (approval ARA 18/27).

Animals and Habituation

Ninety Merino non-pregnant ewes (mean body weight 49.5 kg ± 0.57 kg) comprising 80 test animals and 10 spare animals, aged 7 years, were kept in an animal house and fed standard rations of 200 g blended chaff and 700 g complete pelleted ration (Ridley Agriproducts, Australia; 9.04 MJ/kg dry matter) per animal per day, and provided with water *ad libitum*. The sheep were kept in paddocks prior to the experimental period to allow acclimation to feed and when not being used for training or testing. The experimental protocol is shown in **Figure 1**. The sheep were allocated randomly to one of four ($n = 20$) treatment groups and these were equally divided into four cohorts (5 per treatment, $n = 20$) which were tested on separate days. Cohorts 1 & 2 included 5 spare animals, with the remaining 5 spare animals allocated to cohorts 3 & 4, these spare animals were also allocated to treatment groups and underwent training.

To commence habituation, the first two cohorts were moved into individual pens, under a covered shed which was open on the north. The sheep pens were 2 × 1 m and allowed visual and social

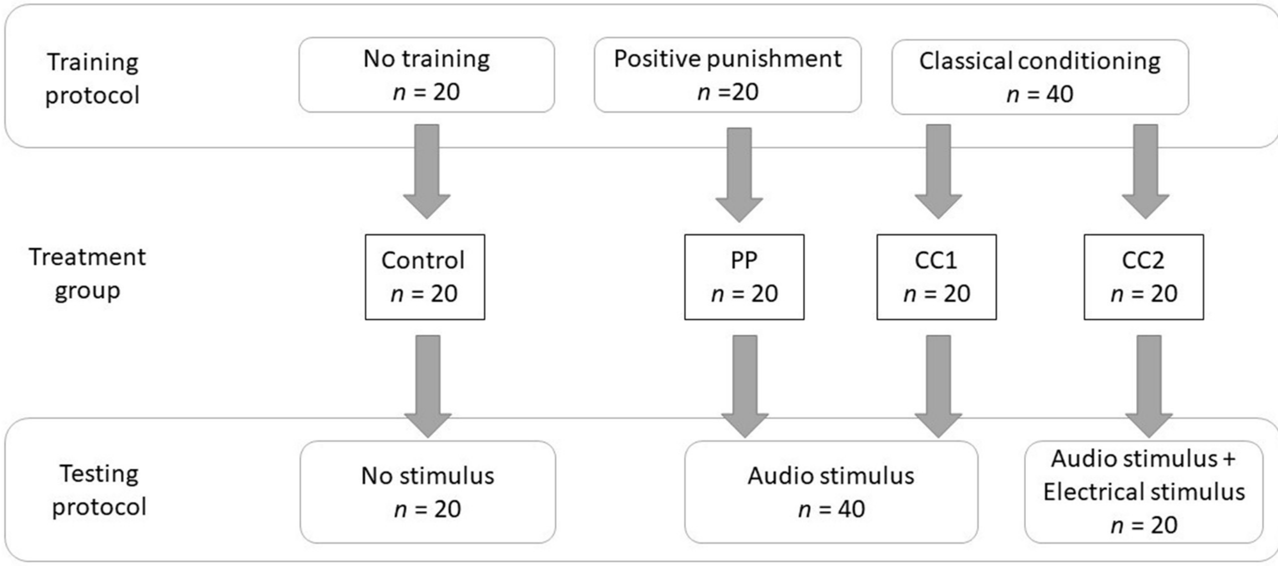


FIGURE 1 | The experimental methodology describing the training and testing protocols of four treatment groups.

interaction. Spare sheep were kept in a larger group pen (3 × 6 m). Training was conducted in laneways adjacent to the animal house facility. All sheep were fitted with dummy collars of similar design and weight to the electronic collars for the duration of the habituation period (14 days). Habituation involved handling and restraining each sheep manually in a standing position for 20 s to simulate blood sample collection, moving to the laneway where they stayed for 1 min and then returning to their pens. All habituation, training and testing of sheep were conducted at similar times of the day. Following the completion of the testing, the first two cohorts of sheep were returned to their paddocks and the third and fourth cohorts were moved into the individual pens to commence training, habituation and testing as described for cohorts one and two. Data collected from two of the sheep were removed from the study, one due to failure to successfully learn the protocol, the other due to inadequate training.

Experimental Design and Treatments

Sheep were randomly allocated to one of four treatments in a randomized design, with each animal being exposed to one treatment only:

1) control—no prior exposure to virtual fencing stimuli and no stimuli applied during testing,
2) audio stimulus after positive punishment training that was predictably controllable (PP),
3) audio stimulus after classical conditioning training that was predictably uncontrollable (CC1), and
4) audio cue and electrical stimulus after classical conditioning training that was predictably uncontrollable (CC2).

Training Protocols

The audio stimulus used in this study was applied remotely from manually controlled dog collars (Garmin TT15, Garmin Ltd., Kansas, KS, USA) at 45–55 dB, 2.7 kHz for a period of 2 s per time. The electrical stimulus was set to level 4 (320 V, 20 μs, 16 pulses per/sec) out of a possible 18. These settings have been utilized in past studies and were effective in achieving successful learning (9, 17).

All sheep except those in the control group underwent training under two distinct protocols: positive punishment and classical conditioning.

Positive Punishment Protocol

The positive punishment treatment was both predictable (audio warning cue given) and controllable (sheep can avoid receiving the shock by responding to the audio cue). The protocol described by Lee et al. (18) aimed to allow training to occur utilizing the behavior of the animal and its responses to the stimuli. Each animal underwent 4 training sessions of 3 min duration each, with one session per day. Previous work in which sheep have been trained in an individual setting have either restricted the number of interactions for welfare reasons (19) or have found that sheep create visual associations and stop interacting with the virtual fence (12, 13). During each session the animal was socially motivated to move through a laneway toward a pen of conspecifics, with the virtual fence located in between (see **Figure 2**). Upon approach to the virtual fence an audio cue was applied using manual controllers operated by experimenters. If the sheep did not stop or turn around, an electrical stimulus was applied. For the PP group average number of electrical stimuli received in the first training session was 3.6 ± 0.46 decreasing to 2.4 ± 0.37 by the second training session, with a maximum of 5 electrical stimuli received in any single training session. An animal was considered to have learned the system when it consistently (two or more times) showed correct responses to the audio cue by either stopping forward

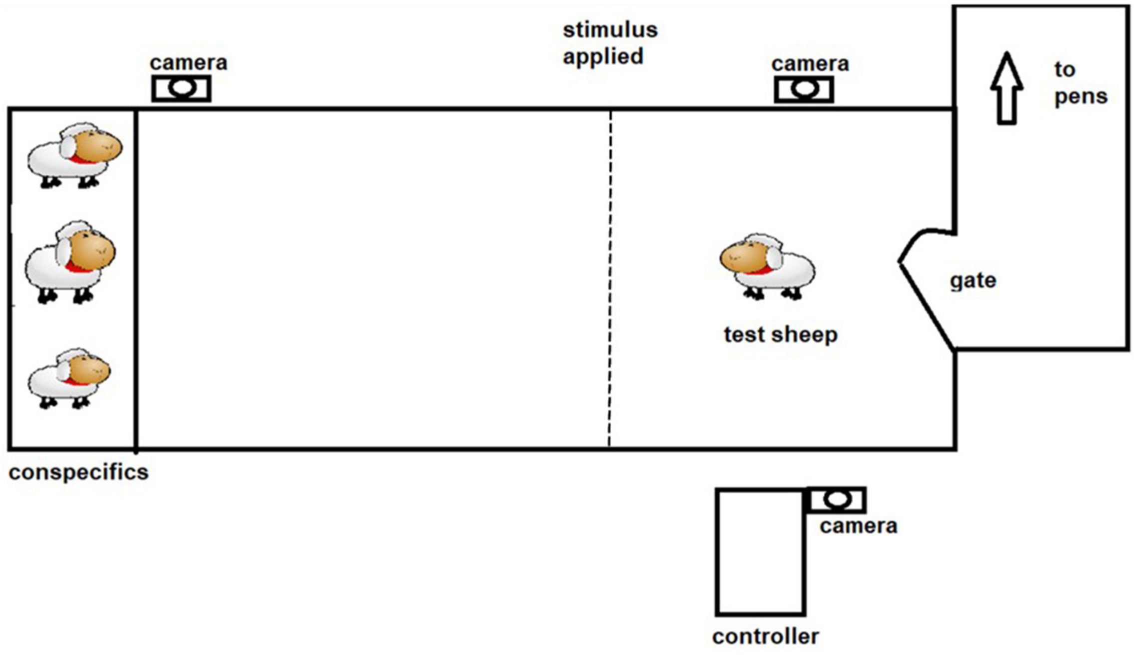


FIGURE 2 | The laneway set up for individual testing and training of sheep.

movement or turning around. One animal failed to successfully learn the system and it was substituted for the test phase with a spare animal.

Classical Conditioning Protocol

The classical conditioning (CC) protocol was predictable but uncontrollable. Each animal underwent 4 training sessions of ∼3 min duration each, with one session per day. During each session the animal was socially motivated to move through a laneway toward a pen of conspecifics, with the virtual fence located in between (see **Figure 2**). Experimenters manually applied the stimuli throughout the training session, irrespective of the behavior exhibited by the animal. Five sequences of the audio (2 s) followed immediately by the electrical stimuli (∼1 s) with 20 s interval in between the sequences were applied per day over 4 consecutive days. Training was suspended early for one animal that showed excessive stress responses, where it attempted to jump out of the laneway.

Testing Stress Responses

Sheep were tested 2 days after the end of their training period, with cohorts one and two tested on consecutive days, and cohorts three and four tested on consecutive days following their training period. Five animals from each treatment were tested individually on each day, totaling 20 animals per treatment over the course of the experiment, and treatment order was randomized for each cohort. Sheep were tested at 5-min intervals, when not being tested they did not have visual or auditory access to the testing arena. For testing, each sheep had their dummy collar removed and replaced with the electronic collar and was moved through a laneway into the test area (∼3 × 15 m). At the end of the test area, a pen holding 3–4 conspecifics served as an attractant. The virtual fencing stimuli were applied immediately upon entry to the test laneway and the test ended after 1 min. The sheep was returned to their pen and their collar was removed.

Body Temperature and Cortisol

Core body temperature is a common measure in the detection of stress in sheep with stress-induced hyperthermia being reported in response to a range of short-term stressors including shearing (20) and isolation (21) and vaginal temperature is a measure of core body temperature (22). As the experiment was conducted during the southern hemisphere summer months, estrus was unlikely to be implicated in body temperature measures. Two days prior to testing, the sheep were fitted with a Thermochron iButton© (Factory calibrated. Model number DS1922L-F5, accuracy 0.5°C, resolution 0.063°C, weight 3.3 g; Maxim International, San Jose, CA, USA) temperature logging device fitted to a intravaginal controlled drug release device previously leached of drug actives (CIDR®, Zoetis, Parsippany, NJ, USA) using polyolefin heat-shrink tubing (23–25). Data were extracted using the program eTemperature version 8.32 (OnSolution, Castle Hill, Australia). Loggers were set to record body temperature in increments of 2-min intervals. The loggers were removed the day after testing. Temperature data was extracted at 10 min before the sheep were restrained for baseline blood sampling and subsequent release into the testing arena (time 0), and at 10, 20, 30, and 60 min following the treatment.

Plasma cortisol is also a commonly used measure in the assessment of welfare in sheep (26). On the test days, each sheep was restrained, and a baseline blood sample (time 0) was collected prior to movement to the test area. All blood samples (10 mL) were taken via jugular venipuncture within 1 min of

TABLE 1 | Ethogram of behaviors measured during the treatment and post-treatment testing periods.

Behavior	Definition
Exploration	Sniffing other sheep, sniffing ground, and sniffing surroundings
Locomotion—stand still	Standing still, all four feet on ground, and stationary
Locomotion—walk	Walking at a slow pace
Locomotion—trot	Medium pace trot
Escape—run	Fast pace run
Turn	Change of direction of at least 90 degrees
Vigilance	Vigilant = head above shoulder; Not vigilant = head parallel to or below shoulder height
Avoidance	Leap with all four feet off the ground, rear with two feet off the ground or fall so that quarters touch the ground, Stretching and rigidity of the neck around the collar, Hunched back posture.
Shake	Shaking head and/or body
Elimination	Urination and/or defecation

restraint and were collected into EDTA coated vacutainer tubes. Additional blood samples were taken at 10, 20, 30, and 60 min following the treatment. Blood samples were centrifuged at 3,000 rpm for 10 min at 4°C on the day of collection, and plasma was retained and stored at −18°C for analysis. Samples were analyzed for plasma cortisol concentration using a commercial radioimmunoassay (Plasma Cortisol RIA, MP Biomedicals, California, CA, USA). This method has been previously validated in our laboratory for use in sheep (27). The intra-assay and inter-assay coefficients of variance (CV) for quality controls containing 24.9, 51.6, and 104.9 nmol/L of cortisol were 5.9, 5.6, and 8.2% and 14.0, 13.3, and 12.5%, respectively.

Behaviors

The behavioral analysis consisted of a number of measures commonly used in sheep welfare analysis, including locomotor activity (28), exploratory behaviors (29), vigilance (30, 31) and avoidance behaviors (9). Video footage was recorded by video camera (Sony Handycam HDR-XR550, Sony Electronics Inc., San Diego, CA, USA), additionally, security cameras were mounted and connected to digital video recorders and captured by IVMS4200 software (Hangzhou Hikvision Digital Technology Co., Ltd). Observations made during testing were recorded and categorized according to the ethogram described in **Table 1** for two measurement periods: The treatment period, lasting 10 s and encompassing the time the treatments were applied; and the remaining 50 s period following the treatment, referred to as "post-treatment." The control treatment was also split into these two measurement periods for equivalence. Locomotion, vigilance and escape behaviors were analyzed as proportion of time spent in the behavior; exploration, turn, avoidance, shake and elimination behaviors were analyzed as count of observations.

Statistical Analysis

All statistical analyses were performed in R (32) using the packages nlme (33), pscl (34), MASS (35), rcompanion (36), dunn.test (37), dplyr (38), and userfriendlyscience (39). Data was tested for normality using visual assessment of Q–Q plots and the Shapiro-Wilk test.

A linear mixed effect model (LMM) with time series was used to analyze cortisol and temperature data. To analyze the cortisol, initial datasets were edited to remove the outliers (two observations from PP and CC2) based on drawn qqplot in R. Cortisol data were log transformed to meet the normality assumptions of LMM in which no more outliers were detected.

Mean ± 2.5 standard deviation (SD) was used to normalize the temperature data which resulted to remove 9 outlier observations [CC1 (1), PP (5, 4 in the same sheep and 1 for another sheep), and CC1 (3, same sheep)] from the dataset. The LMM was used as follows:

$$y_{ijklmn} = \mu + Treatment_i \times Time_j + Cohort_k + \beta_{1l} \times \left(Time1_{il} - \overline{Time1}\right) + Sheep_m + e_{ijklmn}$$

where y_{ijklmn} = response variable (plasma cortisol or temperature at time series point), μ = population mean, $Treatment_i$ = the fixed effect of treatment (4 levels: Control, PP, CC1, CC2), $Time_j$ = the fixed effect of time of measurement (10, 20, 30, and 60 after treatment for cortisol and temperature), $Cohort_k$ = the fixed effect of cohort for cortisol as it was not significant for temperature and eliminated from the model (4 levels: 1, …, 4), $\beta_{1l} \times \left(Time1_{il} - \overline{Time1}\right)$ = the covariate effect of cortisol at time 0 or temperature at time −10, $Sheep_m$ = random effect of sheep, and e_{ijklmn} = random effect of error. To account for the repeated measures over time, a spatial power (since time intervals were not equally spaced) covariance-structure was used in the mixed models for cortisol and body temperature.

A further analysis with an LMM using nlme package was performed in R (32) to investigate the difference between the treatments for within time points. The mathematical model was as follows:

$$y_{ijklm} = \mu + Treatment_i + Cohort_j + \beta_{1k} \times \left(Time1_{1k} - \overline{Time1}\right) + Sheep_l + e_{ijklm}$$

where y_{ijklm} = response variable (plasma cortisol at time 10, 20, 30, and 60 for both plasma cortisol and temperature), μ = population mean, $Treatment_i$ = the fixed effect of treatment (4 levels, control, PP, CC1, CC2), $Cohort_j$ = the fixed effect of cohort for cortisol as it was not significant for temperature and eliminated from the model (4 levels: 1, …, 4), $\beta_{1k} \times \left(Time1_{1k} - \overline{Time1}\right)$ = the covariate effect of cortisol at time 0 or temperature at time −10, $Sheep_l$ = random effect of sheep, and e_{ijklm} = random effect of error. The lsmeans function in the lsmeans package (40) was used to estimate the least square means (LS-means) for all LMMs. The groups were compared using Tukey's test which differences were considered to be significance at $P < 0.05$. The results then plotted using ggplot2 function of R package (32).

Counts of behaviors were separated into the first 10 s during treatment and the 50 s post-treatment. Number of turns were analyzed using a GLM with poisson distribution, the model fitted treatment and day as a fixed effect and the interaction of

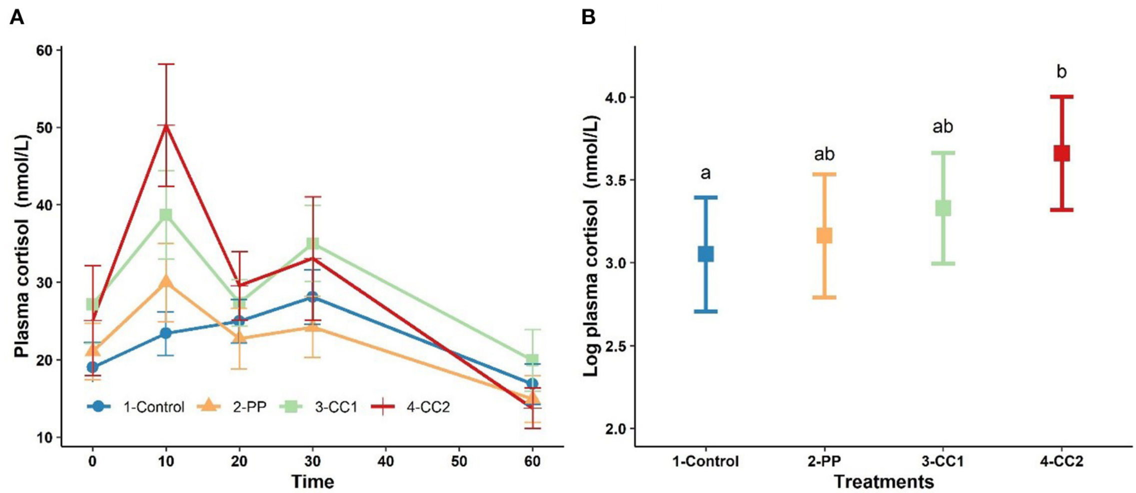


FIGURE 3 | The trend of plasma cortisol changes (mean ± SEM, nmol/L) in response to virtual fencing stimuli on mean over the study time period **(A)** and at time = 10 mins after treatment in sheep. Treatment groups sharing a letter were not significantly different. PP, positive punishment training with audio stimulus in testing; CC1, Classical conditioning training with audio stimulus in testing; CC2, Classical conditioning training with audio and electrical stimulus in testing period. The measures of plasma cortisol in plot **(B)** are based on log transformed data.

treatment and day where appropriate based on ANOVA, QIC and residual deviance of the model. Number of turns in the post-treatment period was over dispersed and required analysis with quasi-poisson distribution. Due to the low occurrence of avoidance, exploration, vocalization, shake and elimination behaviors, these data were placed into a binary frame as either "did" or "did not" perform the behavior. This new data was analyzed using Fishers Exact Tests, examining the number of animals in each group which performed the behaviors. If a significant result was obtained ($P < 0.05$) the data was analyzed *post-hoc* using the package rcompanion (36).

Locomotion data was measured as seconds duration for the treatment period, lasting 10 s, and the post-treatment period, lasting a further 50 s. Data for the treatment observation period could not be transformed to approximate normality, and therefore were subsequently analyzed using a Kruskal-Wallis test followed by Dunn multiple comparison *post-hoc* test with a Bonferroni correction. Stand, trot and run locomotion data for the post-treatment observation period could not be transformed to approximate normality, and therefore were subsequently analyzed using a Kruskal-Wallis test followed by Dunn multiple comparison *post-hoc* test with a Bonferroni correction. Walk data was square root transformed and subsequently was able to meet assumptions of normality (Shapiro-Wilk test) and equal variance (Levene test), this data was then analyzed using a linear mixed effects model with cohort as a fixed effect and individual sheep as a random effect.

RESULTS

Plasma Cortisol

Figure 3A shows the plasma cortisol concentration over time. Cortisol peaked at 10 min for treatments PP, CC1 and CC2. The results from LMM indicated a significant effect of time ($P <$ 0.001), cohort ($P < 0.001$), and the interaction between time × treatment ($P < 0.05$) while treatment was not a significant ($P = 0.32$) factor for plasma cortisol (data not shown). At 10 min, the least square means of plasma cortisol for CC2 treatment was significantly higher than control however, PP and CC1 did not differ ($P > 0.05$) from the control group (**Figure 3B**). At other time points, there were no significant differences between treatments in plasma cortisol (data not shown).

Body Temperature

Body temperature increased over time with a maximum at 30 min after treatment (**Figure 4A**). Based on the obtained results from LMM, time ($P < 0.001$), and the interaction between time × treatment ($P < 0.00$) had significant effects of body temperature, while treatment ($P = 0.13$) and cohort ($P = 0.72$) did not significantly influence body temperature (data not shown). Estimated least squares means for the effect of treatment on body temperature for each time point are presented in **Figures 4B–E**. Body temperature differed between treatments at 10 min with the CC2 treatment having a higher temperature than the control ($P = 0.04$) and PP and CC1 did not differ from any other treatments (**Figure 4B**). At other time points, there were no significant differences between treatments (**Figures 4C–E**) however, the overall differences between treatments tended to be significant at time = 30 ($P = 0.08$) and time- = 60 ($P = 0.06$).

Behavior

Locomotion observations (**Table 2**) showed that during the treatment period, time spent standing still [$H(3) = 16.392$, $P \leq 0.001$], walking [$H(3) = 16.961$, $P \leq 0.001$], and running [$H(3) = 36.491$, $P \leq 0.001$] was significantly different. The CC2 treatment animals exhibited a lower portion of time spent standing than Control ($z = 3.267$, $P_{adj} = 0.007$), PP ($z = 3.583$, $P_{adj} = 0.002$), and CC1 ($z = 2.863$, $P_{adj} = 0.025$) treatments, a lower portion

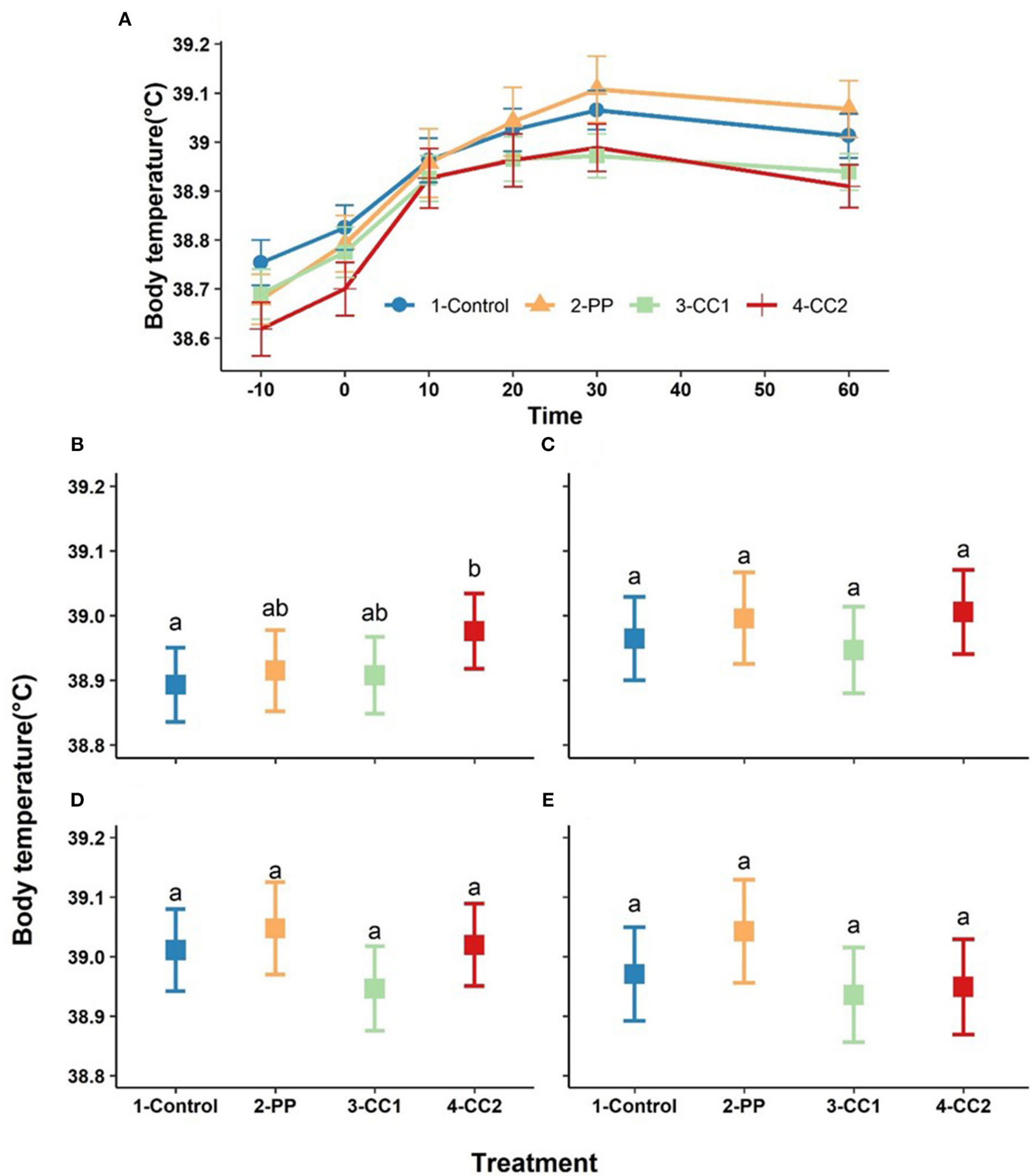


FIGURE 4 | The trend of core temperature changes (mean ± SEM, oC) in response to virtual fencing stimuli on mean over the study time period **(A)** and at the 10, 20, 30, and 60-min post treatment time points **(B–E)** in sheep. Treatment groups sharing a letter were not significantly different within a time point. PP, positive punishment training with audio stimulus in testing; CC1, Classical conditioning training with audio stimulus in testing; CC2, Classical conditioning training with audio and electrical stimulus in testing.

of time walking than the Control ($z = 3.716$, $P_{adj} = 0.001$), and PP ($z = 3.334$, $P_{adj} = 0.005$) treatments; and a longer portion of time running than Control ($z = -5.505$, $P_{adj} \leq 0.001$), PP ($z = -4.782$, $P_{adj} \leq 0.001$), and CC1 ($z = -3.938$, $P_{adj} \leq 0.001$) treatments. There was no significant treatment difference for time spent trotting [$H(3) = 0.820$, $P \leq 0.845$]. For the post-treatment observation period, the time spent standing still [$H(3) = 7.998$, $P = 0.046$], trotting [$H(3) = 17.131$, $P = 0.001$], and running [$H(3) = 28.211$, $P \leq 0.001$] was significantly different. The CC2 treatment animals spent less time standing however there was no significant treatment difference on *post-hoc* multiple comparison analysis. The CC2 animals spent more time trotting than Control ($z = -3.511$, $P_{adj} = 0.003$) and PP ($z = -3.109$, $P_{adj} = 0.011$) treatments, and more time running than Control

TABLE 2 | Locomotion duration in seconds during the treatment period (10 s) and the post-treatment period (50 s).

Behavior	Treatment[1,2] (mean ± SEM)				P-value
	Control ($n = 19$)	PP ($n = 19$)	CC1 ($n = 20$)	CC2 ($n = 20$)	
Stand	2.7 ± 0.49[b]	2.7 ± 0.36[b]	2.4 ± 0.48[b]	0.7 ± 0.26[a]	< 0.001
Walk	4.6 ± 0.76[b]	3.7 ± 0.54[b]	2.9 ± 0.63[a,b]	1.3 ± 0.43[a]	< 0.001
Trot	1.9 ± 0.34[a]	2.4 ± 0.47[a]	2.8 ± 0.58[a]	2.3 ± 0.53[a]	0.845
Run	0.67 ± 0.25[a]	1.1 ± 0.35[a]	1.8 ± 0.46[a]	5.7 ± 0.53[b]	< 0.001
Post-treatment (mean ± SEM)					
Stand	40.1 ± 1.59[a]	40.0 ± 1.64[a]	34.8 ± 2.23[a]	32.9 ± 0.63[a]	0.046
Walk[3]	9.7 ± 1.57[a]	8.9 ± 1.62[a]	10.0 ± 1.44[a]	6.6 ± 1.16[a]	0.241
Trot	0.2 ± 0.12[a]	0.7 ± 0.5[a]	2.5 ± 1.0[a,b]	3.8 ± 1.00[b]	0.001
Run	0.0 ± 0.00[a]	0.5 ± 0.32[a]	1.7 ± 0.88[a]	6.7 ± 2.42[b]	< 0.001

[1]*PP, positive punishment training with audio stimulus in testing; CC1, Classical conditioning training with audio stimulus in testing; CC2, Classical conditioning training with audio and electrical stimulus in testing.*
[2]*For each behavior, means not sharing a common letter within row were statistically different.*
[3]*Post-treatment walk was analyzed by linear mixed effects (LME) model, other behaviors were analyzed by Kruskal Wallis test.*
[a,b]*For each behavior, means not sharing the same superscript within row were not statistically different ($P < 0.05$).*

($z = -4.877$, $P_{adj} \leq 0.001$), PP ($z = -4.200$, $P_{adj} \leq 0.001$), and CC1 ($z = -3.217$, $P_{adj} \leq 0.001$) treatment sheep. There was no significant treatment difference for time spent walking ($F_{3/71} = 1.433$, $P = 0.241$).There was a treatment effect in the number of turns displayed ($P < 0.05$) during treatment, with CC2 animals displaying more turns (mea$n = 1 \pm 0.4$, $P < 0.05$) compared to Control (-2.1, $z = 4.4$), PP (-1.3, $z = -3.9$), and CC1 (-0.7, $z = -2.6$). There was no difference between treatments in the number of turns post-treatment, however a trend was seen between the CC2 and control animals with CC2 animals displaying more turns (mea$n = 1.3$ vs. 0.2 respectively, $t = -1.8$, $P = 0.07$). Behavioral responses to the treatments summarized as did or did not perform are shown in **Table 3**. More animals in the CC2 group displayed avoidance behaviors during treatment compared to the other groups [χ^2 (3) $= 8.2$, $P = 0.02$]. A difference was also seen post-treatment, with fewer CC2 animals displaying exploratory behaviors compared to control and PP [χ^2 (3) $= 13.8$, $P = 0.003$].

TABLE 3 | Behavioral responses to virtual fencing stimuli during treatment (10 s) and post-treatment (50 s) observation periods.

Behavior	Treatment[1,2] (count)			
	Control ($n = 19$)	PP ($n = 19$)	CC1 (n =20)	CC2 ($n = 20$)
Avoidance	1[a]	3[a]	2[a]	11[b]
Exploratory	2	5	6	0
Vocalizations	3	2	2	2
Eliminations	3	5	1	1
Shake	2	4	4	5
Post-treatment (count)				
Avoidance	1	1	2	5
Exploratory	16[a]	17[a]	13[a,b]	8[b]
Vocalizations	7	6	7	4
Eliminations	10	6	13	15
Shake	5	2	0	4

[1]*PP, positive punishment training with audio stimulus in testing; CC1, Classical conditioning training with audio stimulus in testing; CC2, Classical conditioning training with audio and electrical stimulus in testing.*
[2]*For each behavior, means not sharing a common letter within row were statistically different.*
Differing letter superscript [a,b] within row denotes significant difference ($P < 0.05$).
Counts are the total number of animals within the group that displayed the behavior.

DISCUSSION

This study aimed to observe the welfare impact of predictability and controllability of the aversive component of a virtual fence. The sheep which had undergone the predictable controllability (PP) treatment had learned that responding to the audio cue allowed them to control the aversive event, and as expected, we found that the behavioral and physiological responses were not different to the control treatment. This suggests that they perceive this cue as benign once they have learnt how to respond to it. The capacity to predict through an audio warning but not control receiving the aversive stimulus (CC2) induced a higher cortisol and body temperature response compared to the control but was not different to CC1 and PP treatments. However, overall, the inability to control receiving the electrical stimulus (CC2) elicited a stronger behavioral response compared with the other treatments, suggesting that predictability without controllability may be stress inducing. The differences in behavior also suggest that hearing the audio cue (prediction) without receiving the electrical stimulus (CC1) had less impact than hearing the audio cue and receiving the electrical stimulus (CC2), thereby indicating that there is a biological cost to confirmation of uncontrollability.

The plasma cortisol, body temperature and majority of behavioral responses to the audio cue in the animals trained using positive punishment techniques were not significantly different to the control responses, and this is in agreement with earlier work that found the naïve experience of the audio stimulus had no inherent welfare impact (9). This absence of significant differences between the control group and the group trained to the virtual fence using positive punishment suggest that this is a welfare-friendly approach to training sheep to a virtual fence.

The stronger behavioral responses reported in the classically conditioned treatments (CC1 and CC2), particularly increased locomotion, have been linked to stress responses, and may be related to coping strategies (41). In the context of this study, it is likely that most of the running and turning behavior may be explained as an attempt for the animal to escape the situation. It should also be noted that locomotion can increase both cortisol (42) and body temperature responses (43), and may have influenced the stress responses. The importance of controllability in the modulation of the stress response is shown in previous work by Dess et al. (44) in which plasma cortisol responses in dogs exposed to electric shocks were elevated in those dogs which had no control over their experience of the noxious stimuli. Overall, the training protocol using classical conditioning, resulted in increased stress responses and escape behavior, suggesting that the inability to control their exposure to the electrical stimulus was stressful, even if animals were able to predict the aversive event. If this situation were to be on-going, then there would be serious implications for animal welfare, and may result in negative states such as helplessness and hopelessness. These findings should be considered in relation to limitations of the study, including that there was small variation in body temperature in response to the treatments and a small sample size used in the study.

The minimal physiological and behavioral responses observed in the control treatment group indicate that the habituation period was successful in ameliorating stress responses associated with handling and blood sampling which occurred on test days. The observed effect of the treatments on cortisol responses in this study were short-lived, with all sheep returning to baseline within 20 min following the treatment, and behavioral observations reduced in effect from the treatment to the post-treatment observation periods. This is similar to cortisol responses reported in sheep exposed to the acute stress of a barking dog (45). Other previous studies have introduced a stressor for a longer period of time, making appropriate comparison difficult, for example other work exposed sheep to a barking dog for 5 min (46, 47), induced isolation stress for 10 min (48), and longer (49).

In the classic study by Weiss (1) where rats were exposed to electric shocks, animals that had no control over receiving shocks showed a strong stress response (measured by increased corticosteroid levels and the presence of stomach wall lesions). Whereas, when rats were able to prevent receiving an electric shock by turning a wheel, the stress response was not different to controls that did not receive any shocks, indicating that controllability was an important component of the stress response. Interestingly, rats that received a light signal to indicate that a shock was coming (i.e., they were predictable), showed a similar stress response to controls that did not receive any electric shocks. Surprisingly, both the ability to predict and control the occurrence of the electric shocks were equally effective at reducing the stress response, and this was explained by the fact that the animals knew they were experiencing a safe period if they hadn't received a warning signal. In the current study, where the electric shock occurrence in both CC1 and CC2 treatments were predictable (as they were always signaled by an audio cue), but not controllable, the physiological and behavioral stress response was higher in the CC2 treatment compared to controls. As there was no unpredictable uncontrollable treatment, we could not compare the stress response without predictability. The addition of an unpredictable and uncontrollable treatment would be informative, however, this is challenging to test in practice as the test arena/test paradigm itself could become a cue (prediction) for the likelihood of an uncontrollable event occurring. Interestingly, it appears that predictability makes receving an electric shock less aversive. Rats chose predictable electric shock over unpredictable shock, even when the shock duration was up to nine time longer and three times stronger (50). Behaviors are also less disrupted by predictable shock compared with unpredictable shock (51). Further studies to compare predictability with unpredictabilty in the context of the virtual fencing model are recommended.

These findings using virtual fencing as a model begin to provide insights into how predictability and controllability may affect stress responses and animal welfare as proposed in the framework of Lee et al. (8). Another model of relevance to virtual fencing is the Bayesian brain model as described by Colditz (52) in relation to predictive control being linked with physiological stress responses and subsequently affective experience. In this model, the predictions are iteratively refined through the sensory feedback they evoke—i.e., by the potential for the actions to modify and control the sensations. When actions (predictions) aren't able to reduce the discrepancy between expected and actual sensations then the animal becomes stressed. In virtual fencing, once an animal has learned to avoid the fence in response to the audio cue, its situation is both predictable and controllable (for example, the PP treatment), and it can be considered to have agency over its choice to interact with the virtual fence. As demonstrated in this study, the resulting physiological and behavioral stress response to predictable controllability is minimal and thus, we may infer that a negative affective state is not induced due to there being no discrepancy between expected and actual sensations.

CONCLUSIONS

This work highlights the importance of predictability and controllability of events for animal welfare as technology and animal management become more integrated, particularly in systems in which it is necessary for animals to learn in order to be able to be effectively managed.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The animal study was reviewed and approved by CSIRO Chiswick Animal Ethics Committee under the NSW Animal Research Act, 1985 (approval ARA 18/27).

AUTHOR CONTRIBUTIONS

TK, DM, FC, and CL contributed conception and design of the study. TK, SB, DM, and BM conducted the animal experiment. TK and HK performed the statistical analyses. TK wrote the first draft of the manuscript. TK, DM, FC, HK, and CL wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This project was supported by funding through the Australian Government Department of Agriculture, Water and the Environment as part of its Rural R&D for Profit program, the Commonwealth Scientific and Industrial Research Organisation (CSIRO) (internal funding, URL: www.csiro.au/), and the University of New England (School of Environmental and Rural Science project expense support, URL: https://www.une.edu.au/).

ACKNOWLEDGMENTS

Thank you to Troy Kalinowski, Jim Lea, and Tim Dyall (CSIRO) for technical support. A special thank you to Ian Colditz for his feedback and advice on the manuscript. Thanks to Andrew Eichorn (CSIRO) for animal management support and students and support staff from the University of New England and CSIRO for their valuable assistance with this project.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2020.580523/full#supplementary-material

REFERENCES

1. Weiss JM. Psychological factors in stress and disease. *Sci Am.* (1972) 226:104–13. doi: 10.1038/scientificamerican0672-104
2. Greiveldinger L, Veissier I, Boissy A. Emotional experience in sheep: predictability of a sudden event lowers subsequent emotional responses. *Physiol Behav.* (2007) 92:675–83. doi: 10.1016/j.physbeh.2007.05.012
3. Tsuda A, Tanaka M, Hirai H, Pare WP. Effects of coping behavior on gastric lesions in rats as a function of predictability of shock. *J Psych Res.* (1983) 25:9–15.
4. Greiveldinger L, Veissier I, Boissy A. Behavioural and physiological responses of lambs to controllable vs. uncontrollable aversive events. *Psychoneuroendocrinology.* (2009) 34:805–14. doi: 10.1016/j.psyneuen.2008.10.025
5. Destrez A, Deiss V, Leterrier C, Boivin X, Boissy A. Long-term exposure to unpredictable and uncontrollable aversive events alters fearfulness in sheep. *Animal.* (2013) 7:476–84. doi: 10.1017/S1751731112001796
6. Doyle RE, Lee C, Deiss V, Fisher AD, Hinch GN, Boissy A. Measuring judgement bias and emothional reactivity in sheep following long-term exposure to unpredictable and aversive events. *Physiol Behav.* (2011) 102:503–10. doi: 10.1016/j.physbeh.2011.01.001
7. Bassett L, Buchanan-Smith HM. Effects of predictability on the welfare of captive animals. *Appl Anim Behav Sci.* (2007) 102:223–45. doi: 10.1016/j.applanim.2006.05.029
8. Lee C, Colditz IG, Campbell DLM. A framework to assess the impact of new animal management technologies on welfare: a case study of virtual fencing. *Front Vet Sci.* (2018) 5:187. doi: 10.3389/fvets.2018.00187
9. Kearton T, Marini D, Cowley F, Belson S, Lee C. The effect of virtual fencing stimuli on stress responses and behavior in sheep. *Animals.* (2019) 9:30. doi: 10.3390/ani9010030
10. Campbell DLM, Lea JM, Haynes SJ, Farrer WJ, Leigh-Lancaster CJ, Lee C. Virtual fencing of cattle using an automated collar in a feed attractant trial. *Appl Anim Behav Sci.* (2018) 200:71–7. doi: 10.1016/j.applanim.2017.12.002
11. Campbell DLM, Lea JM, Keshavarzi H, Lee C. Virtual fencing is comparable to electric tape fencing for cattle behavior and welfare. *Front Vet Sci.* (2019) 6:445. doi: 10.3389/fvets.2019.00445
12. Marini D, Cowley F, Belson S, Lee C. The importance of an audio cue warning in training sheep to a virtual fence and differences in learning when tested individually or in small groups. *Appl Anim Behav Sci.* (2019) 221:104862. doi: 10.1016/j.applanim.2019.104862
13. Marini D, Meuleman M, Belson S, Rodenburg T, Llewellyn R, Lee C. Developing an ethically acceptable virtual fencing system for sheep. *Animals.* (2018) 8:33. doi: 10.3390/ani8030033
14. Mcgreevy P, Boakes R. Carrots and Sticks: *Principles of Animal Training*. Sydney, NSW: Darlington Press (2011).
15. Blackwell EJ, Twells C, Seawright A, Casey RA. The relationship between training methods and the occurrence of behavior problems, as reported by owners, in a population of domestic dogs. *J Vet Behav.* (2008) 3:207–17. doi: 10.1016/j.jveb.2007.10.008
16. Padalino B, Henshall C, Raidal SL, Knight P, Celi P, Jeffcott L, et al. Investigations into equine transport-related problem behaviors: survey results. *J Equine Vet Sci.* (2017) 48:166–73.e62. doi: 10.1016/j.jevs.2016.07.001
17. Marini D, Llewellyn R, Belson S, Lee C. Controlling within-field sheep movement using virtual fencing. *Animals.* (2018) 8:31. doi: 10.3390/ani8030031
18. Lee C, Prayaga K, Reed M, Henshall J. Methods of training cattle to avoid a location using electrical cues. *Appl Anim Behav Sci.* (2007) 108:229–38. doi: 10.1016/j.applanim.2006.12.003
19. Brunberg EI, Bøe KE, Sørheim KM. Testing a new virtual fencing system on sheep. *Acta Agric Scand A.* (2015) 65:1–8. doi: 10.1080/09064702.2015.1128478
20. Sanger ME, Doyle RE, Hinch GN, Lee C. Sheep exhibit a positive judgement bias and stress-induced hyperthermia following shearing. *Appl Anim Behav Sci.* (2011) 131:94–103. doi: 10.1016/j.applanim.2011.02.001
21. Pedernera-Romano C, Ruiz De La Torre JL, Badiella L, Manteca X. Associations between open-field behaviour and stress-induced hyperthermia in two breeds of sheep. *Anim Welfare.* (2011) 20:339–46.
22. George WD, Godfrey RW, Ketring RC, Vinson MC, Willard ST. Relationship among eye and muzzle temperatures measured using digital infrared thermal imaging and vaginal and rectal temperatures in hair sheep and cattle. *J Anim Sci.* (2014) 92:4949–55. doi: 10.2527/jas.2014-8087
23. Lea JM, Niemeyer DDO, Reed MT, Fisher AD, Ferguson DM. Development and validation of a simple technique for logging body temperature in free-ranging cattle. *Aust J Exp Agric.* (2008) 48:741–5. doi: 10.1071/EA07422
24. Monk JE, Belson S, Colditz IG, Lee C. (2018). Attention Bias Test Differentiates Anxiety and Depression in Sheep. *Front Behav Neurosci.* doi: 10.3389/fnbeh.2018.00246
25. Monk JE, Lee C, Belson S, Colditz IG, Campbell DLM. The influence of pharmacologically-induced affective states on attention bias in sheep. *PeerJ.* (2019) 7:e7033. doi: 10.7717/peerj.7033
26. Niezgoda J, Bobek S, Wronska-Fortuna D, Wierzchos E. Response of sympatho-adrenal axis and adrenal cortex to short-term restraint stress in sheep. *J Vet Med A.* (1993) 40:631–8. doi: 10.1111/j.1439-0442.1993.tb00677.x
27. Paull DR, Lee C, Colditz IG, Atkinson SJ, Fisher AD. The effect of a topical anaesthetic formulation, systemic flunixin and carprofen, singly or in combination, on cortisol and behavioural responses of Merino lambs to mulesing. *Aust Vet J.* (2007) 85:98–106. doi: 10.1111/j.1751-0813.2007.00115.x

28. Verbeek E, Ferguson D, Quinquet de Monjour P, Lee C. Opioid control of behaviour in sheep: Effects of morphine and naloxone on food intake, activity and the affective state. *Appl Anim Behav Sci.* (2012) 142:18–29. doi: 10.1016/j.applanim.2012.09.001
29. Beausoleil NJ, Stafford KJ, Mellor DJ. Sheep show more aversion to a dog than to a human in an arena test. *Appl Anim Behav Sci.* (2005) 91:219–32. doi: 10.1016/j.applanim.2004.10.008
30. Monk JE, Doyle RE, Colditz IG, Belson S, Cronin GM, Lee C. Towards a more practical attention bias test to assess affective state in sheep. *PLoS ONE.* (2018) 13:e0190404. doi: 10.1371/journal.pone.0190404
31. Monk JE, Belson S, Lee C. Pharmacologically-induced stress has minimal impact on judgement and attention biases in sheep. *Sci Rep.* (2019) 9:1–14. doi: 10.1038/s41598-019-47691-7
32. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing (2018).
33. Pinheiro J, Bates D, DebRoy S, Sarkar D. *nlme: Linear and Nonlinear Mixed Effects Models* (2011)
34. Jackman S. *pscl: Classes and Methods for R Developed in the Political Science Computational Laboratory*. Sydney, NSW: United States Studies Centre; University of Sydney (2017).
35. Venables WN, Ripley BD. Modern applied statistics with S. In: Chambers J, Eddy W, Härdle W, Sheather S, Tierney L, editors. *Statistics and Computing*. 4th ed. New York, NY: Springer (2002). p. 12. doi: 10.1007/978-0-387-21706-2
36. Mangiafico S. *rcompanion: Functions to Support Extension Education Program Evaluation_R package* (2018).
37. Dinno A. *dunn.test: Dunn's Test of Multiple Comparisons Using Rank Sums* (2017).
38. Wickham H, François R, Henry L, Müller K. *dplyr: A Grammar of Data Manipulation* (2018).
39. Peters G. *Userfriendlyscience: Quantitative Analysis Made Accessible* (2018).
40. Lenth RV. Least-squares means: the R Package lsmeans. *J Stat Softw.* (2016) 69:1–33. doi: 10.18637/jss.v069.i01
41. Beausoleil NJ. *Physiological Responses of Domestic Sheep (Ovis aries) to the Presence of Humans and Dogs* (Doctor of Philosophy), Massey University, Wellington, New Zealand (2006).
42. Turnbull AV, Rivier CL. Regulation of the hypothalamic-pituitary-adrenal axis by cytokines: actions and mechanisms of action. *Physiol Rev*. (1999) 79:1–71. doi: 10.1152/physrev.1999.79.1.1
43. Brigg P, Pethick DW, Johnson KG, Yovich JV. The influence of wool length on thermoregulation in sheep exercised at different ambient temperatures. In: *Animal Production in Australia*. Wongan Hills: Australian Society Animal Prod (1994). p. 402.
44. Dess NK, Linwick D, Patterson J, Overmier JB, Levine S. Immediate and proactive effects of controllability and predictability on plasma cortisol responses to shocks in dogs. *Behav Neurosci.* (1983) 97:1005–16. doi: 10.1037/0735-7044.97.6.1005
45. Lee TK, Lee C, Bischof R, Lambert GW, Clarke IJ, Henry BA. Stress-induced behavioral and metabolic adaptations lead to an obesity-prone phenotype in ewes with elevated cortisol responses. *Psychoneuroendocrinology.* (2014) 47:166–77. doi: 10.1016/j.psyneuen.2014.05.015
46. Cook CJ. Oxytocin and prolactin suppress cortisol responses to acute stress in both lactating and non-lactating sheep. *J Dairy Res.* (1997) 64:327–39. doi: 10.1017/S0022029997002240
47. Komesaroff PA, Esler M, Clarke IJ, Fullerton MJ, Funder JW. Effects of estrogen and estrous cycle on glucocorticoid and catecholamine responses to stress in sheep. *Am J Physiol.* (1998) 275:E671–8. doi: 10.1152/ajpendo.1998.275.4.E671
48. Caroprese M, Albenzio M, Marzano A, Schena L, Annicchiarico G, Sevi A. Relationship between cortisol response to stress and behavior, immune profile, and production performance of dairy ewes. *J Dairy Sci.* (2010) 93:2395–403. doi: 10.3168/jds.2009-2604
49. Minton J, Apple J, Parsons K, Blecha F. Stress-associated concentrations of plasma cortisol cannot account for reduced lymphocyte function and changes in serum enzymes in lambs exposed to restraint and isolation stress. *J Anim Sci.* (1995) 73:812–7. doi: 10.2527/1995.733812x
50. Badia P, Culbertson S, Harsh J. Choice of longer or stronger signalled shock over shorter or weaker unsignalled shock. *J Exp Anal Behav*. (1973) 19:25–32. doi: 10.1901/jeab.1973.19-25
51. Davis H, Levine S. Predictability, control, and the pituitary-adrenal response in rats. *J Comp Physiol Psychol.* (1982) 96:393–404. doi: 10.1037/h0077892
52. Colditz IG. Objecthood, agency and mutualism in valenced farm animal environments. *Animals.* (2018) 8:50. doi: 10.3390/ani8040050

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 reviewer DB declared a past co-authorship with one of the authors CL to the handling editor.

frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 06 January 2021
doi: 10.3389/fvets.2020.600846

Suitability of Slower Growing Commercial Turkey Strains for Organic Husbandry in Terms of Animal Welfare and Performance

Anna Olschewsky[1]*, *Katharina Riehn*[2] *and Ute Knierim*[1]

[1] *Farm Animal Behavior and Husbandry Section, Faculty of Organic Agricultural Sciences, University of Kassel, Witzenhausen, Germany,* [2] *Department Ecotrophology, Faculty of Life Sciences, Hamburg University of Applied Sciences, Hamburg, Germany*

Intensive turkey production with fast growing strains is often critically discussed regarding animal welfare problems. Studies evaluating the welfare status of both organic and less intensive selected turkey strains are limited, except in the slightly slower growing Kelly Broad Breast Bronze (Kelly). The aim of this study was to assess the welfare of turkeys from two strains with further decreased growth rate, Hockenhull Large Bronze (HoBr) and Hockenhull Black (HoBl), in comparison to Kelly under commercial organic conditions with 100% organic feed. Altogether 844 non-beak-trimmed male turkeys (274–288 per line) were reared and fattened in three replications with each six groups. On group level, use of resources in the 7, 16, and 25th week of life, mortality and feed conversion were recorded. Each bird was assessed with regard to plumage and skin condition as indicators of agonistic interactions, cannibalism and feather pecking, with regard to leg health, footpad, breast skin condition and, as performance indicators, live and carcass weight, utilization, daily weight gain and weights of valuable meat parts. The significantly slower growing HoBl showed slightly fewer malposition of the legs, reduced injury rates and less breast buttons, but a higher susceptibility to footpad dermatitis than Kelly turkeys. HoBr with a similar growth rate compared to Kelly had slightly more problems concerning walking ability and plumage damage, but also less breast buttons than Kelly turkeys. However, effect sizes were negligible ($\Phi < 0.10$), except for the higher occurrence of footpad dermatitis and the reduced number of breast buttons in HoBl with small effect sizes ($\Phi = 0.20$–0.24). Use of resources, prevalence of breast blisters and mortality, were not statistically different, although mortality rate was numerically lower in HoBl. Thus, for none of the studied strains clear benefits or disadvantages in terms of the birds' predisposition for welfare problems could be identified. Overall, prevalences of animal welfare problems were mostly lower than in comparable studies while performances were comparatively high. Therefore, turkeys from the studied strains appear to be suitable for organic rearing and fattening with 100% organic feed, given a good management.

Keywords: animal welfare, behavior, health, performance, turkey strains

Edited by:
Alan G. McElligott,
City University of Hong Kong,
Hong Kong

Reviewed by:
Jessica Elisabeth Martin,
University of Edinburgh,
United Kingdom
Oluwaseun Serah Iyasere,
Federal University of Agriculture,
Abeokuta, Nigeria
Shawna Weimer,
University of Maryland, College Park,
United States

***Correspondence:**
Anna Olschewsky
olschewsky@uni-kassel.de

Specialty section:
This article was submitted to Animal Behavior and Welfare, a section of the journal Frontiers in Veterinary Science

Received: *31 August 2020*
Accepted: *07 December 2020*
Published: *06 January 2021*

Citation:
Olschewsky A, Riehn K and Knierim U (2021) Suitability of Slower Growing Commercial Turkey Strains for Organic Husbandry in Terms of Animal Welfare and Performance. Front. Vet. Sci. 7:600846. doi: 10.3389/fvets.2020.600846

INTRODUCTION

Turkey meat production is increasing worldwide (1). The EU countries produce around 33% of the world's turkey meat (~ 2 million tons in 2018) and are therefore the largest producer following the USA (2). In 2019, Germany was the largest producer in Europe with almost 500,000 tons of turkey meat (2). Most of these turkeys are reared and fattened under conventional conditions, while < 2% of the housing places are certified organic (3). In the European Union, organic animal husbandry is regulated by EU-Regulation 889/2008 (4). Among others, the rules shall safeguard higher animal welfare standards, based e.g. on higher space allowances or access to a free-range (4). This regulation further stipulates: "To prevent the use of intensive rearing methods, poultry shall either be reared until they reach a minimum age or else shall come from slow-growing poultry strains". Since the middle of the 20th century turkey breeding is economically very successful, not at least because of intensive methods of cross-breeding (5). As a result, the live weight of turkeys has quadrupled compared to wild turkeys (6). The use of fast growing turkeys for meat production dominates in Germany as well as in the rest of the world (7).

However, turkey husbandry is often critically discussed because of several potential animal welfare problems. Among them is a high risk for skin injuries and damaged or lost feathers due to agonistic interactions and the behavioral disorders cannibalism and feather pecking (5). Dalton et al. (8) summarized agonistic interactions, cannibalism and feather pecking as injurious pecking. Extensive injuries particularly on top of the head and large feather losses can lead to the death of animals (8). Causes are multifactorial, including genetic, nutritional and husbandry factors (8, 9). Rather similar ranges of affected birds have been reported from investigations of commercial and experimental, conventional and organic conditions, with 77–100% of turkeys with plumage damage (10–12) as well as 23–39% with injuries (10–15). Further animal welfare problems may be respiratory and cardiovascular diseases (5) as well as high prevalences of footpad dermatitis, impaired leg health and breast skin alterations. Again, different European and North American studies of mainly fast, but also slower growing turkeys under commercial, experimental, conventional or organic conditions have found similar ranges of these welfare problems. Footpad dermatitis was present in more than 80% of investigated birds (14, 16–21). Footpad dermatitis can appear in various forms of hyperkeratosis and necrosis due to inflammatory processes (22, 23) and is caused or promoted by a number of factors, with litter moisture playing a decisive role (24, 25). Impaired leg health with reduced walking ability or malposition of the legs can be linked to a multitude of genetic, nutritional and husbandry factors (5, 26). Prevalences of more than 50% of turkeys with reduced walking ability or malposition of the legs were found (10, 27, 28). In addition, breast skin alterations, which are also indicative of impaired welfare (29–31), were present in 8–48% of birds (10, 12, 21, 32–36). These alterations include both breast buttons, which are lesions in unfeathered areas, as well as breast blisters, which are inflammations of the *bursa sternalis*, with breast buttons being the more frequent findings (10, 34, 35).

It is supposed that problems regarding health or behavior in turkeys are mainly attributed to the high growth potential of the commonly used commercial hybrids and the intensive husbandry conditions in conventional farming (5, 36). However, the outcomes in organic husbandry are not fundamentally different. One reason for this may be the widespread use of conventional fast growing turkeys in organic husbandry (3) due to the EU regulation 889/2008 (4) stipulating minimum slaughter ages that are compatible with common strains suitable for cutting, since there is hardly any demand for whole turkeys in Germany (3, 7). Some organic farms, nevertheless, keep a slightly slower growing commercial turkey line, Kelly Broad Breast Bronze (BBB) (3), for which slightly reduced prevalences of welfare problems have been reported (10, 28, 34, 37). The majority of German organic farmers is keeping female turkeys of fast-growing strains, mostly B.U.T Big 6 (3), because of their sex-specific lower nutritional demands, and the lower growth rate which may contribute to a reduced susceptibility to welfare problems (36). This is e.g. reflected by reports of lower prevalences of damaging feather pecking (13) and breast skin alterations (11, 21, 29). However, all the mentioned welfare outcomes are still not satisfactory, considering the high welfare level expected in organic farming (4). Furthermore, it is debatable whether the use of predominantly one sex in organic farming is an acceptable practice. In our opinion the assessment of the suitability of a strain for organic farming should be based on the most challenging condition which is the rearing and fattening of male turkeys with 100% organic feeding.

It was therefore the aim of this study to assess the welfare of male turkeys under organic husbandry conditions in two strains with further decreased growth rate in comparison to Kelly BBB.

MATERIALS AND METHODS

Birds and Husbandry

Three fattening batches (cycles) from July 2015 to January 2018 on a commercial organic farm in Northern Germany were monitored. Rearing and fattening conformed to the Animal welfare - farm animal husbandry ordinance (38), the EU regulation 889/2008 for organic farming (4) as well as to Demeter guidelines (39). Management (climate control, health care, animal controls) was in accordance with standard commercial guidelines (40).

Besides the reference Kelly BBB (Kelly), the strains Hockenhull Large Bronze (HoBr, recommended for free-range husbandry, nearly similar growth potential as Kelly BBB) and Hockenhull Black (HoBl, recommended as robust, lower target live weight, markedly reduced growth potential), both from Aviagen (UK), were used. They had been selected based on enquiries in Germany, UK and France in 2015 considering the criteria (a) lower, but sufficient growth potential, (b) stated robustness, (c) suitability for cutting, and (d) commercial availability. Hockenhull turkeys were delivered from the UK, and

TABLE 1 | Average nutrient contents in the rations including supplemented wheat in the different feeding phases over three fattening batches (cycles).

	Ration 1	Ration 2	Ration 3	Ration 4	Ration 5	Ration 6
Week of life	1–3	4–9	10–13	14–15	16–19	20–25
ME-poultry (MJ/kg)	11.4	11.8	11.3	11.4	9.0	7.9
Raw protein (%)	23.9	22.3	19.4	17.5	15.5	13.6
Methionine (%)	0.44	0.43	0.34	0.3	0.3	0.2
Lysine (%)	1.21	1.16	0.89	0.8	0.7	0.6
Sodium (%)	0.21	0.19	0.16	<0.02	<0.02	<0.02
Starch (%)	29.8	31.0	34.5	31.1	27.6	24.2
Fiber (%)	5.0	5.2	6.7	6.0	5.4	4.7

Kelly turkeys from the Netherlands by car as non-beak-trimmed, male 1-day-poults. The aim was to rear and fatten 100 turkeys per strain and batch, divided in two groups of 50 individuals. This was repeated twice, so that three batches were conducted. Due to deviations in delivered numbers and sex of birds, altogether 844 male turkeys were investigated instead of the planned 900 birds. The birds were individually marked at the end of rearing, using colored and numbered leg bands. These were changed in the 13th week of life in order to adapt to the birds' growth.

Due to the seasonal availability of HoBl, each batch started at the end of July or beginning of August. Rearing (1–6th week of life) and fattening (7–25th week of life) took place in a mobile house which contained six pens (each measuring 3.5 × 5.0 m) and a central control aisle. The birds had access to separate winter gardens (each 2.5 × 5.0 m) from the 5th week of life onwards. These were attached to the mobile house and had a roof but no solid floor which was, however, littered. On the sides they were covered with wood in the lower part and with windbreak netting in the upper part. The pens inside the house and the winter gardens were separated by wire mesh fences so that the groups could not mix. At the end of the rearing phase the turkeys received access to separate grass-covered free-range areas (500 m² per group) with electric fencing. Each pen was equipped with feeders (small, plastic round feeders in the rearing period and bigger, metal ones in the fattening period, all filled by hand) and round drinkers (1st—3rd week: bell drinkers filled by hand, 4th—25th week: plasson drinkers connected to the regular water supply). Their height was regularly adapted to the growth of the turkeys. In the 2nd week of life, round, wooden perches (Ø 2 cm) were placed in each pen. With the start of the fattening period, they were replaced by larger perches (round metal perches, Ø 4 cm, in the first batch and wooden perches, squared timbers of 3 × 6 cm, in the second and third batch). Additionally, sand was provided inside the pens during the rearing phase, which was replaced by a mixture of grit and sea shells as well as pecking blocks at the beginning of the fattening period. Pens were littered mainly with wood shavings from day 1 until the 25th week of life every couple of days as needed. In the fattening period also straw was used. The house was supplied with regular electricity, ventilation was semi-automatic, and gas radiant heaters provided heating.

Compound feed with 100% organic components was purchased and fed in six feeding phases (**Table 1**). From the 14th week of life onwards the compound feed was supplemented with increasing proportions of the farm's own wheat (wheat %: 14–15th week = 10%, 16–19th week = 20%, 20–25th week = 30%).

Conforming to the farm's usual practice, the birds were slaughtered consecutively in the 17, 20, and 25th week of life, respectively.

Recording of Behavioral, Health and Performance Measures

In the 7, 16, and 25th week of life all animals of every group inside the pen that used either the "feeding area" or "perches" were counted via instantaneous scan sampling (41) with a 15-min-interval. Observations were carried out by one person on four consecutive days during each time 4 h (i.e., covering the light period from 9 A.M. to 5 P.M. twice) and noted in check sheets. In parallel, birds using the winter garden were counted in the same way and time intervals (scans) based on videos. Finally, the number of birds using the free-range area was calculated as the rest of birds from the total number minus those recorded in the pen and winter garden. Number of recordings were partly reduced due to free-range closure because of histomonosis infections, a camera failure and removal of perches to prevent animal accidents. Intra- and inter-observer reliability for video recordings and inter-observer-reliability between two persons for direct observations were both good ($r = 0.89$–1.0, $n = 17$ observations of each 2 h).

After the end of the behavioral observations, the physical condition of all birds was assessed. The applied clinical scoring schemes were adapted from earlier studies in order to allow maximal comparability of results (**Table 2**). In the 7, 16, and 25th week of life, the assessment of walking ability and leg position was conducted: all individuals were carefully encouraged to walk along a path of 3 m for the gait scoring. Then they jumped on a straw bale where their leg position was rated. If a bird refused to move or showed poor locomotion, they were not forced to walk the entire 3 m distance and to jump on the bale. Afterwards they were fixed in upright position by one person on the straw bale

TABLE 2 | Scoring schemes, modified after the given references, with definitions, and results of inter-assessor-reliability testing for the measures regarding physical condition.

Score	Plumage (without tail feathers)	Skin	Gait	Leg position	Footpads	Breast blister	Breast button
0	Completely intact and smooth plumage (including tips of feathers), no bare skin areas[A]	No injuries, up to 3 point-like bruises	Upright, steady striding, toes are bent backwards when the leg is lifted	parallel	Intact skin, no swelling	No breast blister	No breast button
1	Single feathers damaged (pecked, disheveled or broken), no bare skin areas[A]	Superficial, point-like injuries or >3 bruises or bloodied feather follicles	Slight abnormality, foot is quickly put down again after lifting, toes are not bent	x-shaped, smaller distance between hocks than between feet	Hyperceratosis, small necrotic spots or slight swelling concerning footpads or toes	Small round swelling, can be fluctuating	Point-like hardening
2	Several feathers damaged or bare skin areas[A] ≤2.5 cm (largest diameter)	Deeper and larger injuries ≤2.5 cm (largest diameter)	Strong pendulum locomotion due to lameness on one or both sides	Wide-legged, the legs are parallel but with a wider distance (leg position at rump at outer side)	Larger necrotic areas <50% of footpad or at ≤2 toes or moderately swollen toes	Fist-sized fluctuating or hardened swelling	Skin lesion ≤2 cm (largest diameter)
3	Many feathers damaged or bare skin areas[A] >2.5 −10 cm (largest diameter)	Deep and large injuries >2.5 cm (largest diameter)	Bird sits down again as soon as possible or can only move with great effort (e.g., flapping of wings)	o-shaped, greater distance between hocks than between feet	Large necrotic areas >50% of footpad or at >2 toes or severely swollen toes	Double fist-sized fluctuating or hardened swelling	Skin lesion >2 cm (largest diameter) ≥ 2 cm
4	Large areas with damaged feathers or bare skin[A] >10 cm (largest diameter)	–	–	–	–	–	–
References	(10–12, 33)	(33)	(10, 12)	(10, 12)	(42)	(10)	(10)
Inter-assessor-reliability (PABAK)	K = 0.73–0.90 (n = 90)	0.69–1.0 (n = 120)	0.90–1.0 (n = 60)	0.60–1.0 (n = 90)	0.81–0.87 (n = 111)	0.87–0.90 (n = 30)	0.87–0.90 (n = 30)

[A] *Including bare areas with feather follicles visible.*

TABLE 3 | Mortality and causes over three fattening batches (cycles).

Causes of death	Kelly BBB		Hockenhull bronze		Hockenhull black	
	n	%	*n*	%	*n*	%
First 2 weeks of life: unthrifty birds or leg deformities	7	2.4	6	2.1	6	2.1
Accidents	0	0.0	4	1.4	5	1.7
Histomonosis	10	3.4	7	2.5	2	0.7
Cannibalism	1	0.3	2	0.7	0	0.0
Enteritis	1	0.3	2	0.7	1	0.4
Hepatitis	1	0.3	1	0.4	0	0.0
Unknown cause	1	0.3	3	1.1	1	0.4
Total losses	21		25		15	
Total mortality	7.2		8.8		5.2	

or a table (depending on the bird's weight) for the investigation of plumage, skin condition and footpads (by consecutively lifting legs backwards). Their plumage condition at the back, sides and wings and skin injuries at the whole body, except the breast, were scored while stroking back the feathers. For the assessment of the footpads in the 7 and 16th week of life, both footpads were cleaned with a brush and additionally with water, if necessary, before scoring. In case of different findings on both footpads, the worse result was documented. In addition, footpads of all slaughtered animals were examined in the 20 and 25th week of life in the same way. The occurrence of breast blisters and breast buttons was recorded after slaughter and plucking in all birds.

Furthermore, losses were documented, and carcasses subjected to pathological-anatomical examinations by the Veterinary Pathology Department of the University of Leipzig. Also, the results of the official *ante-mortem* and *post-mortem* slaughter-inspections were documented.

Regarding performance, all living turkeys were weighed with a manual poultry scale (BAT1, VEIT electronics with a capacity of max. 50 kg and an accuracy of 1 g) in the 7, 16, and 25th week of life, and additionally, birds intended for slaughter in the 17 and 20th week of life. Individual carcass weights were taken 4 h after evisceration (without head and legs) using a digital scale. Utilization was calculated as percentage of carcass from live weight. In addition, the weights of valuable meat parts (breast, upper and lower legs) were determined from 10% of slaughtered and cutted turkeys. The individual daily weight gains were calculated for the birds slaughtered in the 25th week of life. Feed conversion for each group was calculated based on the provided feed that had been continuously recorded. Back-weighing of feed was conducted each before weighing of the turkeys. Losses of feed inside the pens were not recorded.

For all measures based on scoring, an acceptable to good inter-assessor-reliability between two persons was ascertained (**Table 2**).

Statistical Analysis

Data on group level, such as the behavioral data, were analyzed using non-parametric tests due to small sample sizes: Friedmann test was used to ascertain possible time effects over the three observation points of time, independent of treatments (strain), which were compared using Kruskal-Wallis test. Kruskal-Wallis test was also used to analyze possible treatment effects on mortality and feed conversion.

Individual data per animal, in case of ordinal data, were dichotomized by converting score 0 to 0 and score > 0 to 1. Generalized linear mixed models (in r, package lme4 v1.1-21, glmer with glmerControl optimized by Bobyqa) were applied with the fixed factors "strain," "week of life" and their interaction, and the random effect "animal nested in group and batch" for repeated measurements, or "group nested in batch" for single measurements. Using the package lsmeans v2.30-0 mean and confidence interval averaged over the levels of "week of life" were estimated for binomial data. The reference strain Kelly was defined as intercept in order to allow comparisons with the two other strains. Metric individual data were analyzed in the same way, with the same factors, with linear mixed models (package lme4 v1.1-21, lmer with lmerControl, optimized by Bobyqa and lmerTest v3.1-3). Normal distribution of residuals as well as variance homogeneity were checked in r (QQ-normal plot, skewness, kurtosis and scatter plot) and, in case of live and carcass weight, utilization and lower leg weight, data transformed in r by logarithmizing the square root. Nevertheless, normal distribution could not be reached for the variables live weight, carcass weight and utilization. Therefore, Kruskal-Wallis test and *post-hoc* Mann-Whitney-*U*-test were used. The script for the mentioned calculations in r are listed in **Supplementary File 1**.

In addition, effect sizes were calculated using SPSS for all test results. According to Ellis (43), phi-correlation (Φ) was computed for dichotomized data and point-biserial correlation (r_{pb}) for metric data, respectively.

RESULTS

Mortality rates did not differ significantly between the strains ($p = 0.41$, $\chi^2 = 1.77$, df $= 2$, $n = 6$, $r_{pb} = 0.16$–0.46). The causes of losses and in some cases necessary cullings of animals are displayed in **Table 3**.

Histomonosis was diagnosed by histological and PCR examinations of tissue samples from caecum and liver in all three batches and led to altogether 19 losses (**Table 3**). However, during the official *post-mortem* slaughter inspections only a few pathologic-anatomical changes were found in the surviving animals.

Losses due to cannibalism were recorded once in Kelly and twice in HoBr (**Table 3**). Kelly had significantly more injuries than HoBl, but not than HoBr (**Table 4**). The majority of injuries was superficial and point-like (score 1). Score 2 was found in 0.4–15.0% and score 3 in 0.3–10.0% of assessments (**Figure 1**). Overall, a significant effect of the factor "week of life" was detected ($p < 0.01$) but no significant interaction between "strain" and "week of life."

Regarding plumage condition, in all turkeys and at all assessment times predominantly slight damage of single feathers (score 1) was found. Score 4 was never present, score 3 only once and score 2 in 0.8–4.0% of all cases (**Figure 2**). Significantly less HoBr turkeys had a completely intact plumage than Kelly, while HoBl did not differ (**Table 4**). However, there were significant interactions between "week of life" and "strain": The proportion

TABLE 4 | Results of generalized linear mixed models and effect sizes (phi-correlation) regarding possible effects of "strain" and "week of life" on the turkeys' physical condition including estimated mean and confidence interval averaged over the levels of "week of life"; significant interactions are stated in the text (Kelly = Kelly Broad Breast Bronze, HoBr = Hockenhull Bronze, HoBl = Hockenhull Black).

Measure	Line	Mean yes (%) CI	Est.	Std. error	*z*-value	*p*-value	OR	Φ
Skin injuries (yes vs. no)	Kelly vs.	30.2 16.4; 48.8						
	HoBr	31.0 16.9; 49.7	0.035	0.184	0.191	0.848	1.05	0.01
	HoBl	22.2 11.4; 38.7	−0.414	0.187	−2.216	0.027	0.78	−0.06
Plumage damage (yes vs. no)	Kelly vs.	88.5 83.9; 91.9						
	HoBr	92.9 89.0; 95.4	1.150	0.541	2.125	0.034	1.49	0.07
	HoBl	90.3 86.3; 93.3	−0.053	0.394	−0.135	0.893	1.55	0.08
Impaired walking ability (yes vs. no)	Kelly vs.	1.0 0.4; 2.6						
	HoBr	3.0 13.3; 7.0	1.132	0.480	2.360	0.018	2.80	0.10
	HoBl	0.9 0.3; 2.3	−0.115	0.517	−0.222	0.825	1.03	0.00
Malpositions of legs (yes vs. no)	Kelly vs.	10.5 6.5; 16.3						
	HoBr	8.4 5.1; 13.6	−0.244	0.302	−0.805	0.421	0.85	−0.03
	HoBl	5.3 3.1; 9.0	−0.733	0.312	−2.347	0.019	0.55	−0.09
Footpad dermatitis (yes vs. no)	Kelly vs.	24.3 8.0; 54.1						
	HoBr	35.6 13.1; 66.9	0.344	0.448	0.767	0.443	1.56	0.11
	HoBl	64.6 33.4; 87.0	1.390	0.439	3.170	0.002	2.68	0.24
Breast blister (yes vs. no)	Kelly vs.	2.7 0.8; 8.5						
	HoBr	2.2 0.6; 7.1	−0.244	0.508	−0.480	0.631	0.77	−0.02
	HoBl	1.4 0.4; 5.3	−0.663	0.562	−1.181	0.238	0.51	−0.05
Breast buttons (yes vs. no)	Kelly vs.	9.8 6.2; 15.1						
	HoBr	3.4 1.7; 6.8	−1.134	0.402	−2.819	0.005	0.33	−0.13
	HoBl	1.1 0.3; 3.4	−2.319	0.620	−3.739	<0.001	0.10	−0.20

CI, 95% confidence interval; Est., Estimate; Std. error, Standard error; OR, Odds ratio; Φ, Phi-correlation.

of turkeys with no plumage damage decreased in Kelly between the 16 and 25th week of life, while it increased in HoBl ($p < 0.01$) and changed little in HoBr ($p = 0.04$). Additionally, a significant effect of "week of life" was detected ($p < 0.01$).

Time effects on the use of the different resources were identified regarding the feeding area ($p < 0.01$, $\chi^2 = 16.78$, df = 2, $n = 18$), perches ($p < 0.01$, $\chi^2 = 16.89$, df = 2, $n = 12$) and winter gardens ($p < 0.01$, $\chi^2 = 28.78$, df = 2, $n = 18$), but not for the free-range area ($p = 0.51$, $\chi^2 = 1.33$, df = 2, $n = 6$; **Figure 3**). No significant differences between the strains were found (analyzed separately for each observation period, except for the free-range area) ($p = 0.06$–0.90, $\chi^2 = 0.22$–5.55, df = 2, $n = 4$–6; **Figure 3**).

Except for Hockenhull Bronze in the 25th week of life, walking ability was mostly not impaired, but from the 16th week of life onwards score 3 was found once in each strain and score 2 in

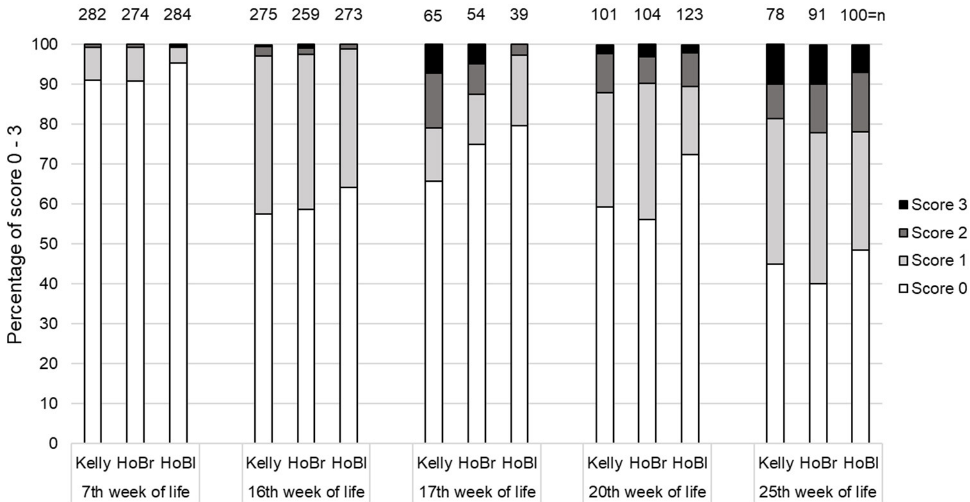


FIGURE 1 | Percentages of turkeys with different skin injury scores (score 0 = intact, score 1 = superficial spots, score 2 < 2.5 cm, score 3 > 2.5 cm) at five scoring times, in three strains (Kelly = Kelly BBB, HoBr = Hockenhull Bronze, HoBl = Hockenhull Black) over three fattening batches (cycles), each with two groups per strain, with n = total number of individuals assessed.

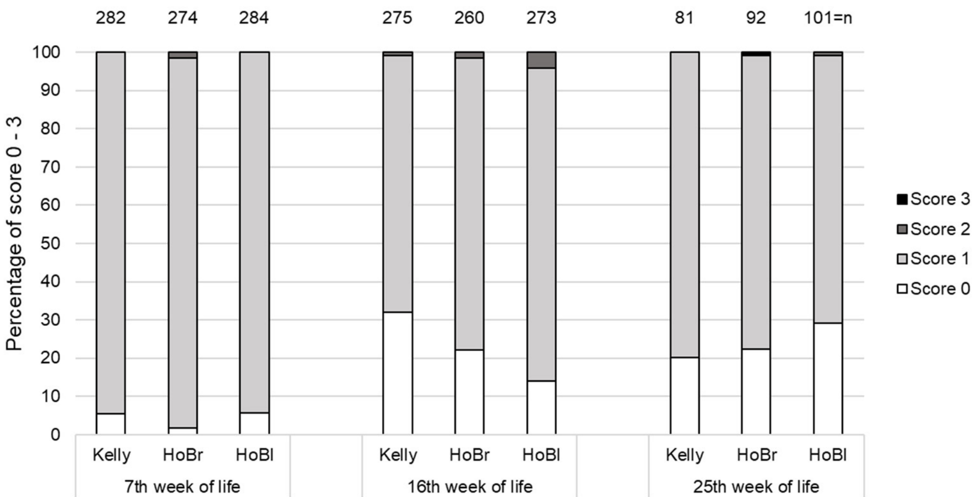


FIGURE 2 | Percentage of turkeys with different plumage condition scores (score 0 = intact, score 1 = single feathers pecked or broken, score 2 = several feathers pecked, broken or smaller bare skin areas, score 3 = larger plumage damage or larger bare skin areas) at three scoring times in three strains (Kelly = Kelly BBB, HoBr = Hockenhull Bronze, HoBl = Hockenhull Black) over three fattening batches (cycles), each with two groups per strain, with n = total number of individuals assessed.

0.8% (HoBl) and 5% of birds (HoBr). Thus, turkeys with impaired walking ability predominantly showed only slight abnormalities (score 1; **Figure 4**). Kelly was significantly less affected than HoBr, but not than HoBl (**Table 4**). No significant interaction between "strain" and "week of life" was found, whereas a significant effect of the latter was detected ($p < 0.01$).

Significantly more abnormal leg positions were detected in Kelly turkeys ($n = 638$) than in HoBl ($n = 659$), but not than in HoBr ($n = 626$; **Table 4**). They comprised mainly x-shaped legs and no o-shaped legs (definitions see **Table 2**). Some malpositions were already visible in the 7th week, with a marked increase until the 25th week of life. Overall, a significant effect of the factor "week of life" was detected ($p < 0.01$) but no significant interaction between "strain" and "week of life."

Footpad dermatitis affected Kelly significantly less than HoBl, but not than HoBr (**Table 4**), although HoBr showed larger increases of prevalences from the 7 to 16th week of life compared to Kelly (interaction: $p = 0.02$). Furthermore, a significant effect of "week of life" was detected ($p < 0.01$). Most alterations of the footpad were minor or medium (score 1 and score 2), from 7 to 57%, and only a small proportion (0–5%) were inflammations of larger areas of the footpad or toes (score 3; **Figure 5**).

Breast blisters occurred in 0–6% ($n = 686$) of slaughtered turkeys per strain and assessment time, with no significant

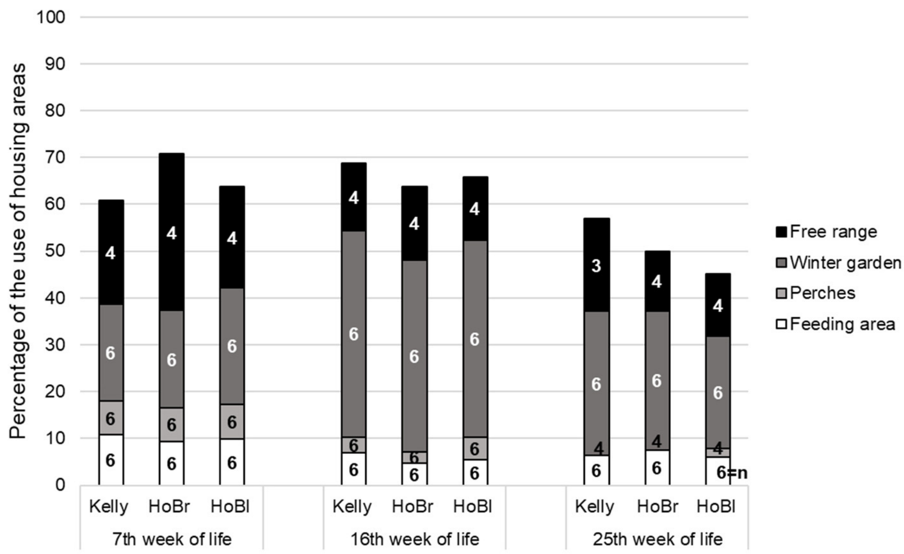


FIGURE 3 | Use of different resources during eight light hours (average percentage of observed time per animal) at three observation time points in three strains (Kelly = Kelly BBB, HoBr = Hockenhull Bronze, HoBl = Hockenhull Black) with $n = 3$ fattening batches (cycles) * 2 groups (partly reduced n: free-range closure due to histomonosis infection, camera failure, removal of perches due to animal accidents).

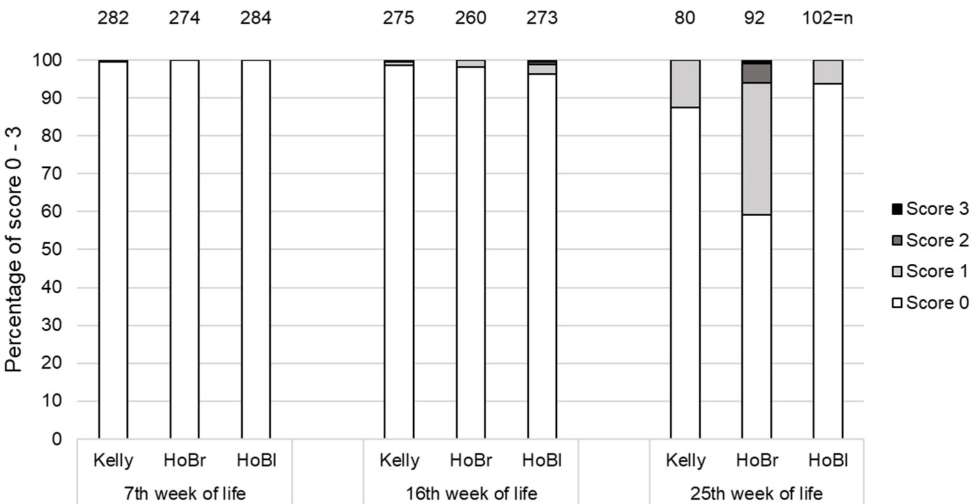


FIGURE 4 | Percentages of turkeys with different gait scores (score 0 = normal, score 1 = slight abnormality, score 2 = defined lameness, score 3 = unable to walk) at three scoring times in three strains (Kelly = Kelly BBB, HoBr = Hockenhull Bronze, HoBl = Hockenhull Black) over three fattening batches (cycles), each with two groups per strain, n = total number of individuals assessed.

differences between strains (**Table 4**). Breast buttons were more prevalent and occurred significantly more in Kelly ($n = 244$) than in HoBr ($n = 247$) and HoBl ($n = 260$; **Table 4**). For both alterations no significant interactions between "strain" and "week of life" or significant effects of the latter was found ($p = 0.20$–0.88). In total, only four birds (0.9–2%) (HoBr and Kelly) had breast buttons with a diameter of more than 2 cm (score 3). Up to 11% of Kelly turkeys showed smaller lesions (score 2), and the rest (0.9–5% of all lines) had very small lesions.

Concerning performance measures, Kelly achieved at all assessment times significantly higher average live weights (21.4 kg, 25th week) than HoBr (19.5 kg, 25th week) and HoBl (16.6 kg, 25th week; **Figure 6**; **Table 5**).

Furthermore, the average carcass weight was significantly higher in Kelly (16.6 kg, 25th week) than in HoBr (15.3 kg, 25th week) and HoBl (12.8 kg, 25th week) at most assessment times (**Table 5**). Utilization (Kelly and HoBl: 75%, HoBr: 76%) only differed at 20th week between Kelly and HoBl (**Table 5**). Feed

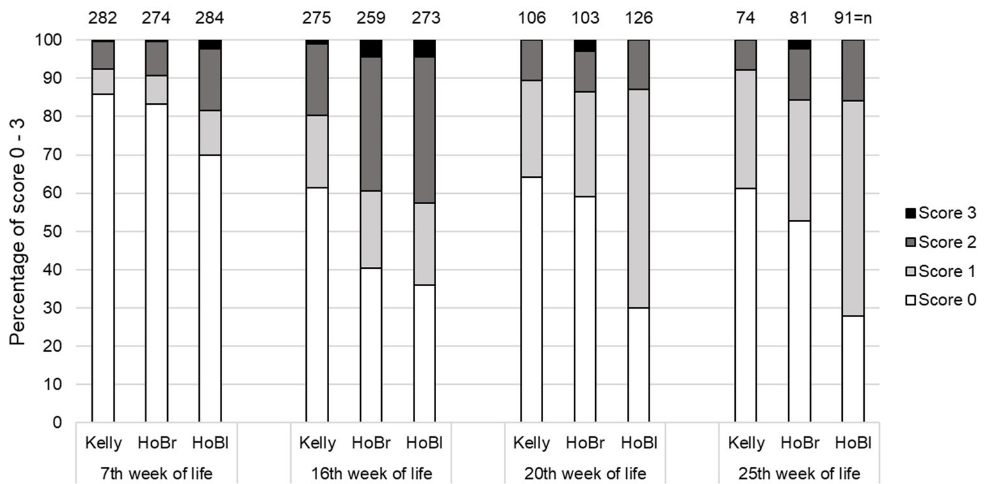


FIGURE 5 | Percentages of turkeys with different footpad scores (score 0 = intact, score 1 = small necrotic spots, score 2 < 50% necrotic footpad, score 3 > 50% necrotic footpad) at four scoring times in three strains (Kelly = Kelly BBB, HoBr = Hockenhull Bronze, HoBl = Hockenhull Black) over three fattening batches (cycles), each with two groups per strain, n = total number of individuals assessed.

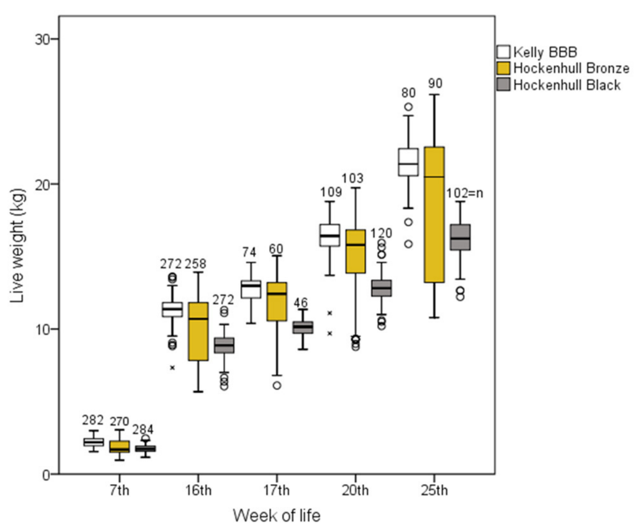


FIGURE 6 | Live weight of three strains at five weighing times from three fattening batches (cycles), each with two groups per strain, with n (total number of weighed individuals) given; box plot with median, lower and upper quartile, minimum and maximum and outliers.

TABLE 5 | Results of Kruskal-Wallis, Mann-Whitney-*U*-test and effect sizes (point-biserial correlation) regarding possible effects of "strain" on turkey performance in the 25th week (Kelly = Kelly Broad Breast Bronze, HoBr = Hockenhull Bronze, HoBl = Hockenhull Black).

	Week	Kruskal-Wallis test				Mann-Whitney-*U*-test		
		n	χ^2	df	*p*-value		*p*-value	r_{pb}
Live weight	25	80–102	92.43	2	<0.01	Kelly vs. HoBr	0.01	−0.35
						Kelly vs. HoBl	<0.01	−0.87
Carcass weight	25	72–90	78.64	2	<0.01	Kelly vs. HoBr	0.04	−0.30
						Kelly vs. HoBl	<0.01	−0.85
Utilization	25	49–67	1.41	2	0.49	–	–	–
						–	–	–

df, degrees of freedom; r_{pb}, point-biserial correlation.

conversion did not differ between strains (Kelly: 2.9:1, HoBr: 2.8:1, HoBl: 3.1:1 kg feed:kg live weight; $p = 0.60$, $\chi^2 = 1.02$, $n = 5–6$, reduced n due to lost documentation).

Average daily weight gain from day 1 until the 25th week of life was significantly higher in Kelly with 111 g than in HoBl with 84 g, but not statistically different from HoBr with 101 g (**Table 6**).

Including all assessment times, the average breast weight after slaughter was significantly higher in Kelly (5.9 kg, 25th week) than in HoBl (4.7 kg, 25th week), but not compared to HoBr (5.7 kg, 25th week; **Table 6**). The same applied to the weight of the upper leg where Kelly reached 2.8 kg, HoBl 2.0 kg and HoBr 2.7 kg in the 25th week (**Table 6**). The weight of the lower leg was altogether higher in Kelly (2.1 kg, 25th week) than both in HoBl (1.4 kg, 25th week) and HoBr (1.9 kg, 25th week; **Table 6**). However, there were significant interactions between "week of life" and "strain": HoBl showed a reduced increase in weight of upper and lower leg at the end of fattening period ($p = 0.01–0.03$). In case of breast weight, no significant interaction between "week of life" and "strain" were found, whereas a significant effect of "week of life" was found for breast weight, upper and lower leg weight ($p < 0.01$).

DISCUSSION

The expectation that strains differ in the extent of welfare problems depending on their growth potential (5, 36) was not

TABLE 6 | Results of linear mixed models and effect sizes (point-biserial correlation) regarding possible effects of "strain" and "week of life"* on performance measures; significant interactions are stated in the text (HoBr = Hockenhull Bronze, HoBl = Hockenhull Black).

Parameter	Line	Est.	Std. error	df	*z*-value	*p*-value	r_{pb}
Daily weight gain	Kelly vs. HoBr	−9.911	6.702	13.079	−1.479	0.163	−0.35
	Kelly vs. HoBl	−26.432	6.695	13.023	−3.948	0.002	−0.88
n	Kelly = 80, HoBr = 102, HoBl = 90						
Breast meat	Kelly vs. HoBr	−0.355	0.272	12.914	−1.306	0.214	−0.25
	Kelly vs. HoBl	−1.022	0.273	13.134	−3.743	0.002	−0.67
n	Kelly = 30, HoBr = 32, HoBl = 30						
Upper leg	Kelly vs. HoBr	−0.170	0.170	32.384	−0.997	0.326	−0.12
	Kelly vs. HoBl	−0.387	0.178	36.510	−2.175	0.036	−0.52
n	Kelly = 30, HoBr = 31, HoBl = 30						
Lower leg	Kelly vs. HoBr	−0.259	0.118	28.049	−2.201	0.036	−0.14
	Kelly vs. HoBl	−0.454	0.122	31.598	−3.705	0.001	−0.39
n	Kelly = 31, HoBr = 31, HoBl = 30						

*For analysis of daily weight gain a possible effect of "week of life" was not included.
Est., Estimate; Std. error, Standard error; df, degrees of freedom; r_{pb}, point-biserial correlation.

confirmed in this study for the range of growth rates investigated. In general, almost all differences in animal welfare outcomes between the strains were of negligible ($\Phi < 0.10$) to small effect size ($\Phi < 0.20$–024). In contrast, other studies found strain differences, mostly comparing fast vs. slower growing strains, for cannibalism and feather pecking (10, 13, 14, 28) and partially for leg health (10, 44) vs. (28, 45) as well as breast skin health (10) vs. (21). While in the present study no statistical strain differences in terms of breast blisters were found, the slowest growing HoBl showed reduced prevalences of breast buttons with a small effect size ($\Phi = 0.20$). HoBl additionally showed slightly fewer malpositions of the legs and reduced injury rates, but on the other hand had more cases of footpad dermatitis (again with small effect size: $\Phi = 0.24$), for which a strain effect has not been reported before (14, 16, 19, 21, 45). Similarly, mixed results were found for HoBr turkeys that showed a rather similar growth rate compared to Kelly. They had slightly more problems concerning walking ability and plumage damage, but also less breast buttons than Kelly turkeys. Further, no statistical strain differences could be detected regarding the use of resources, contrary to the majority of earlier studies (10, 32, 44, 46) vs. (19, 28), and regarding mortality rate. Although it was numerically lower in HoBl, the variation between groups and batches was high and sample size on group level low. Thus, none of the studied strains showed clear benefits or disadvantages in terms of the birds' predisposition for welfare problems.

Concerning the general welfare and performance level of the monitored birds, the average mortality rates from 5.2 to 7.2% per strain during rearing and fattening can be regarded moderate compared to other study results (organic husbandry with different strains and not always including the rearing period) that ranged from 2 to 21% (10, 19, 47–49). In particular, considering the *Histomonas meleagridis* infection with commonly high mortalities of about 90–100% (50), it may even be deemed low. This may partly be explained by the immediate measures that were taken to limit the effects of the infection and that included closure of the free-range (3 weeks in batch one, 3 days in batch two and three), covering the litter with corrugated board and applying quicklime around the house.

Only three birds died as a result of cannibalism, which in all cases happened after single animals entered another group by jumping over the fence in the free-range area. Also, Buchwalder and Huber-Eichler (51) described that turkeys attack unknown animals more likely than animals from their own group. Despite lacking systematic behavioral observations of injurious pecking, it is very likely that apart from these very specific cases, no cannibalism occurred. Nevertheless, about one third of birds had injuries, mostly of superficial and small extent, which according to chance observations were caused by agonistic interactions between the males. This is an important difference to domestic fowl where injuries due to agonistic interactions can almost only be found around the combs. For turkeys, however, it is not clear whether injuries reported in the literature are related to cannibalism or agonistic behavior (10, 11, 13, 14, 21, 36). Only Spindler (15) explicitly states that in her study around 34% of the recorded injuries were due to cannibalism. Savory (52) described that often conspicuous bloody spots or damaged skin areas can trigger cannibalistic behavior. Therefore, it is noteworthy, that in our study, despite the presence of injuries from fighting, no outbreaks of cannibalism occurred. Comparisons of injury levels with other studies are further hampered by their partly lacking reports of the proportion of the different injury scores (13, 36, 49). However, the majority of researchers found roughly similar prevalences of injuries and also of plumage damage (10–15).

Similarly, although the majority of birds showed slight damage of single feathers, and a few birds had single lacking feathers, there was no indication of a manifest feather pecking problem in the monitored groups. Most of the damage was likely mechanically caused, because birds often came into contact with the equipment of the comparatively small pens or the pen partitions. In fact, this was the reason why in the last batch, the perches were removed in the last 3 weeks. In this context it is important to note that the study conditions were different from usual commercial husbandry conditions with commonly much larger groups than 50 birds and consequently larger absolute space allowances. On the other hand, it is possible that the small group sizes, especially during the rearing phase, contributed to the lack of cannibalism and feather pecking problems. This should be further investigated.

The use of perches during the day was, with around 4% of the observed time, generally low compared to other investigated slower growing turkeys [(32): 12% and 16%; (46): 10–31%]. The latter results only included young turkeys until the 12th week of life, and it is known that use of perches decreases with increasing live weight (46). Moreover, the observation time in comparable studies covered longer periods (32, 46). In this study, natural illumination levels in the winter garden were too low in the early morning and in the evening hours for reliable observations. A contribution to reduced perch use was probably the use of metal perches in the first batch; the acceptance by the animals was low. Additionally, the winter gardens, which were not available in the studies cited above, were intensively used (31%) and provided no perches. Furthermore, Berk et al. (12) found a lower use of winter gardens of about 9–11%. The free-range area, on the other hand, was less used (19%) then reported elsewhere [(10, 28, 32, 44): 36–94%], with the exception of Straßmeier (28) who found a reduced use in winter with 7–12%. According to Bergmann (10) and Straßmeier (28) the use of the free-range decreases with falling temperature. Since the access to the free-range in our study was mainly provided in fall and winter, this, together with the attractive winter gardens, might explain the lower use. No comparable figures regarding use of the feeding area are available. The time budget of 13–48% for foraging behavior, reported by Bircher and Schlup (53) for commercial lines, is higher than the 7% found here, but foraging behavior also occurs at other places than at the feeder in the pen where it was only recorded.

Leg health (walking ability, position of legs) was comparatively good in our study, considering reports of more than 50% of the turkeys having problems in this area (10, 27, 28). Our results may already reflect the recently increased breeding efforts to improve animal health and in particular leg health (54). Furthermore, Bergmann (10) found better leg health in winter. Thus, a seasonal effect might have contributed to the comparatively good results. Interestingly, the slightly worse walking ability of HoBr corresponded with a lower weight of the lower legs compared to Kelly despite rather similar live weights. On this line, Nestor et al. (55) found increased shank widths related to better walking ability.

Also footpad lesions were a less frequent problem (36–59% of birds) than in the majority of comparable studies that reported prevalences of more than 80% (14, 16–21). Still, there is room for improvement also regarding the results of the present study. The higher affliction of HoBl with the lowest live weights is in line with conclusions from Habig et al. (20) that genetic factors other than live weight influence footpad health. Alterations of the breast skin (3–10% of birds) were similarly or less prevalent compared to other studies with reported prevalences of 8–48% (10, 12, 21, 32–36).

Performance in general was high, considering the 100% organic diet fed from day 1 onwards. In other studies under conventional or organic conditions, Kelly reached mostly lower live and carcass weights (10, 34, 37, 47), whereas live weights of B.U.T Big 6 in comparable studies exceeded these results most of the time (10, 19, 34). However, in comparison breast weight of Kelly was lower (37). It is possible that the relatively low contents of methionine in the ration might have played a role.

CONCLUSIONS

Despite partly differences in growth rate and predominantly slight differences in predispositions for welfare problems, no clear advantage or disadvantage of a specific strain could be identified. Overall, prevalences of animal welfare problems were mostly lower than in comparable studies and predominantly consisted of only minor alterations. Therefore, all monitored turkey strains appear to be suitable for rearing and fattening under organic conditions with 100% organic feed, given a good management, in terms of performance and animal welfare. However, it should be emphasized that group sizes were smaller than under usual commercial conditions. Therefore, it would be useful to conduct further investigations in larger groups to verify the results.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The study was carried out in accordance with the German animal protection act (56). The protocol for the clinical monitoring and behavioral observations was reported to the competent authority (Nds. Landesamt für Verbraucherschutz und Lebensmittelsicherheit, LAVES) without objection and followed the "Guidelines for ethical treatment of animals in applied animal behavior and welfare research" (57).

AUTHOR CONTRIBUTIONS

AO designed the project together with UK (recording in live animals) and KR (recording in slaughtered animals). AO conducted data collection, analyses, and wrote the first draft of the manuscript. UK assisted in data analysis and interpretation of results. All authors contributed to the preparation of the manuscript and approved the submitted version.

FUNDING

The project was supported by funds of the Landwirtschaftliche Rentenbank and the Mahle-Stiftung.

ACKNOWLEDGMENTS

Special thanks to Christine Bremer and the whole Team of the farm Bauckhof. We also wish to thank the Veterinary Pathology Department of the University of Leipzig for pathological-anatomical examinations and advice concerning Histomonosis detection. For the latter we are also grateful to the Department of Poultry Diseases of the University of Berlin, Prof. Dr. Hafez. Dr. Eike Rommelfanger helped with the statistical analysis. Finally, we thank the teams of UK and KR who supported the project in a number of ways.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2020.600846/full#supplementary-material

Supplementary File 1 | Script for the calculations in *r*.

REFERENCES

1. Windhorst H. The dynamics of turkey meat production and trade. In: Hafez HM, editor. *Turkey Production: Toward Better Welfare and Health: Proceedings of the 5th International Symposium on Turkey production 28.−30.5.2009*. Berlin: Mensch-und-Buch-Verlag (2009). p. 152–68.
2. FAOSTAT. Livestock Primary: Production of Meat, turkey in World (2018). Available online at: http://www.fao.org/faostat/en/#data/QL/visualize
3. Schaack D, Quaing H, Nusch T, Rampold C, Beck MM. *Schlussbericht Analyse des Bio-Geflügelmarktes*. (2018). Available online at: https://orgprints.org/33738/1/33738-15OE071-ami-schaack-2018-biogefluegelmarkt.pdf
4. Verordnung (EG) Nr. 889/2008 der Kommission vom 5. September 2008 mit Durchführungsvorschriften zur Verordnung (EG) Nr. 834/2007 des Rates über die ökologische /biologische Produktion und Kennzeichnung von ökologischen/biologischen Erzeugnissen hinsichtlich der ökologischen/biologischen Produktion, Kennzeichnung und Kontrolle: EG-Öko-Verordnung (2008).
5. Hafez HM. Übersicht über Probleme der haltungs- und zuchtbedingten Erkrankungen bei Mastputen. *Archiv Geflügelkunde*. (1996) 60:249–56.
6. Hünigen H, Mainzer K, Hirschberg RM, Custodis P, Gemeinhardt O, Al Masri S, et al. Structure and age-dependent development of the turkey liver: a comparative study of a highly selected meat-type and a wild-type turkey line. *Poult Sci.* (2016) 95:901–11. doi: 10.3382/ps/pev358
7. Meyer H. Putenherkünfte – Übersicht zum aktuellen Leistungsstand. *Bericht Kartzfehn*. (2000) 67:1–6.
8. Dalton HA, Wood BJ, Torrey S. Injurious pecking in domestic turkeys: development, causes, and potential solutions. *World's Poultry Sci J.* (2013) 69:865–76. doi: 10.1017/S004393391300086X
9. Marchewka J, Watanabe TT, Ferrante V, Estevez I. Review of the social and environmental factors affecting the behavior and welfare of turkeys (meleagris gallopavo). *Poult Sci.* (2013) 92:1467–73. doi: 10.3382/ps.2012-02943
10. Bergmann SM. *Vergleichende Untersuchung von Mastputenhybriden (BUT. Big 6) und einer Robustrasse (Kelly Bronze) bezüglich Verhalten, Gesundheit und Leistung in Freilandhaltung*. (Dissertation). München (2006).
11. Bartels T, Böhme J, Cramer K, Dellavolpe A, Ellerich R, Ludewig M, et al. *Abschlussbericht zum Forschungsauftrag 06HS015, Indikatoren einer tiergerechten Mastputenhaltung*. (2009). Available online at: https://www.ml.niedersachsen.de/download/72908/Abschlussbericht_zum_Forschungsauftrag_06HS015_Indikatoren_einer_tiergerechten_Mastputenhaltung_.pdf
12. Berk J, Hinz T, Wartemann S. Tierverhalten, Tierleistung und Tiergesundheit in einem Putenmaststall mit Aussenklimabereich. *Landbauforschung Völkenrode*. (2006) 56:159–71.
13. Busayi RM, Channing CE, Hocking PM. Comparisons of damaging feather pecking and time budgets in male and female turkeys of a traditional breed and a genetically selected male line. *Appl Anim Behav Sci.* (2006) 96:281–92. doi: 10.1016/j.applanim.2005.06.007
14. Große Liesner BB. *Vergleichende Untersuchungen zur Mast- und Schlachtleistung sowie zum Auftreten (Häufigkeit/Intensität) primär nicht-infektiöser Gesundheitsstörungen bei Puten fünf verschiedener Linien*. (Dissertation). Hannover (2007).
15. Spindler B. *Pathologisch-anatomische und histologische Untersuchungen an Gelenken und Fußballen bei Puten der Linie BUT. Big 6 bei der Haltung mit und ohne Außenklimabereich*. (Dissertation). Hannover (2007).
16. Bellof G, Brandl M, Schmidt E. *Schlussbericht Forschungsprojekt Nr. 03OE234 – Ökologische Putenmast: Abstimmung von Genotyp, Haltung und Fütterung* (2010). Available online at: http://orgprints.org/18771/1/18771-06OE234-hswt-bellof-2010-oekologischePutenmast.pdf
17. Krautwald-Junghanns M-E, Ellerich R, Mitterer-Istyagin H, Ludewig M, Fehlhaber K, Schuster E, et al. Examinations on the prevalence of footpad lesions and breast skin lesions in British United Turkeys Big 6 fattening turkeys in Germany. Part I: prevalence of footpad lesions. *Poult Sci.* (2011) 90:555–60. doi: 10.3382/ps.2010-01046
18. Allain V, Huonnic D, Rouina M, Michel V. Prevalence of skin lesions in turkeys at slaughter. *Br Poult Sci.* (2013) 54:33–41. doi: 10.1080/00071668.2013.764397
19. Berk J, Schumacher C, Krautwald-Junghanns ME, Martin M, Bartels T. Verweildauer von Mastputen verschiedener Herkünfte im Bereich von Tränke- und Fütterungseinrichtungen. *Landbauforschung Appl Agric For Res.* (2013) 63:245–54. doi: 10.3220/LBF_2013_245-254
20. Habig C, Spindler B, Beyerbach B, Kemper N. Evaluation of footpad health and live weights in two lines of turkey hens kept under organic husbandry conditions in Germany. *Berliner Münchener Tierärztliche Wochenschrift*. (2017) 130:250–7. doi: 10.2376/0005-9366-15111
21. Krautwald-Junghanns ME, Bartels T, Berk J, Deerberg F, Dressel A, Erhard MH, et al. *Abschlussbericht zum Forschungsauftrag Indikatoren einer tiergerechten Mastputenhaltung unter den Bedingungen der ökologischen Geflügelmast* (2017) Available online at: http://orgprints.org/34305/1/34305-12OE030-uni-leipzig-krautwald_junghanns-2017-mastputen.pdf
22. Martland MF. Wet litter as a cause of plantar pododermatitis, leading to foot ulceration and lameness in fattening turkeys. *Avian Pathol.* (1984) 13:241–52. doi: 10.1080/03079458408418528
23. Platt S. *Die reticulate scales an den Fußballen schwerer Mastputen und deren Beeinflussung durch unterschiedliche Biotindosierungen unter Feldbedingungen*. (Dissertation). Berlin (2004).
24. Mayne RK. A review of the aetiology and possible causative factors of foot pad dermatitis in growing turkeys and broilers. *World's Poultry Sci J.* (2005) 61:256–67. doi: 10.1079/WPS200458
25. Abd El-Wahab A, Visscher CF, Beineke A, Beyerbach M, Kamphues J. Experimental studies on the effects of different litter moisture contents and exposure time to wet litter on development and severity of foot pad dermatitis in young fattening turkeys. *Archiv für Geflügelkunde*. (2012) 76:55–62.
26. Ferket P. Leg problems in turkeys. In: Hafez HM, editor. *Turkey Production: Toward better Welfare and Health: Proceedings of the 5th International Symposium on Turkey production 28.−30.5.2009*. Berlin: Mensch-und-Buch-Verlag (2009). p. 113–23.

27. Wartemann S. *Tierverhalten und Stallluftqualität in einem Putenmaststall mit Aussenklimabereich unter Berücksichtigung von Tiergesundheit, Leistungsmerkmalen und Wirtschaftlichkeit*. (Dissertation). Braunschweig (2005).
28. Straßmeier P. *Einfluss von Strukturelementen, Futterzusammensetzung und Witterung auf das Verhalten von gemischt gehaltenen Big Six und Kelly Bronze Puten in der Auslaufhaltung*. (Dissertation). München (2007).
29. Gonder E, Barnes HJ. Focal ulcerative dermatitis ("breast buttons") in marketed turkeys. *Avian Dis.* (1987) 31:52–8.
30. Kamyab A. Enlarged sternal bursa and focal ulcerative dermatitis in male turkeys. *World's Poultry Sci J.* (2001) 57:5–12. doi: 10.1079/WPS20010002
31. Berk J. *Artgerechte Mastputenhaltung: Baulich-technische Ansätze zur Verbesserung der Haltungsumwelt*. Darmstadt: KTBL (2002). p. 81.
32. Berk J, Hahn G. Aspects of animal behaviour and product quality of fattening turkeys influenced by modified husbandry. *Archiv Tierzucht Dummerstorf.* (2000) 43:189–95.
33. Ellerbrock S. *Beurteilung verschiedener Besatzdichten in der intensiven Putenmast unter besonderer Berücksichtigung ethologischer und gesundheitlicher Aspekte*. (Dissertation). Hannover (2000).
34. Schweizer CH. *Gesundheit, Leistung und Fleischqualität von gemischt gehaltenen B.U.T big 6 und Kelly Bronze Puten in der Auslaufhaltung*. (Dissertation). München (2009).
35. Mitterer-Istyagin H, Ludewig M, Bartels T, Krautwald-Junghanns M-E, Ellerich R, Schuster E, et al. Examinations on the prevalence of footpad lesions and breast skin lesions in B.U.T. Big 6 fattening turkeys in Germany. Part II: prevalence of breast skin lesions (breast buttons and breast blisters). *Poult Sci.* (2011) 90:775–80. doi: 10.3382/ps.2010-01142
36. Ermakow O. *Ergebnisse der Fleischuntersuchung bei Puten aus ökologischer und konventioneller Haltung*. (Dissertation). Leipzig (2012).
37. Le Bris J. *Gesundheit, Leistung und Verhalten konventioneller Mastputenhybriden unter den Bedingungen ökologischer Haltungsanforderungen*. (Dissertation). München (2005).
38. Verordnung zum Schutz landwirtschaftlicher Nutztiere und anderer zur Erzeugung tierischer Produkte gehaltener Tiere bei ihrer Haltung (Tierschutz-Nutztierhaltungsverordnung): TierSchNutztV (2017).
39. Demeter EV. *Richtlinien 2020: Erzeugung und Verarbeitung Richtlinien für die Zertifizierung ≫Demeter≪ und ≫Biodynamisch≪* (2020). Available online at: https://www.demeter.de/leistungen/zertifizierung/richtlinien
40. Verband Deutscher Putenerzeuger EV. *Bundeseinheitliche Eckwerte für eine freiwillige Vereinbarung zur Haltung von Mastputen*. (2013). Available online at: https://www.bmel.de/SharedDocs/Downloads/DE/_Tiere/Tierschutz/ZDG-Eckwerte-Haltung-Mastputen.pdf?__blob=publicationFile&v=5.
41. Martin P, Bateson P. Recording methods. In: Martin P, Bateson P, editors. *Measuring Behaviour: An Introductory Guide*. Cambridge, NY: Cambridge University Press (2007). p. 48–61.
42. Verband Deutscher Putenerzeuger e.V. Pododermatitis-Scoring. (2014).
43. Ellis PD. *The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation of Research Results*. Cambridge: Cambridge Univ Press. (2011). p. 173.
44. Cottin E. *Einfluss von angereicherter Haltungsumwelt und Herkunft auf Leistung, Verhalten, Gefiederzustand, Beinstellung, Lauffähigkeit und Tibiale Dyschondroplasie bei männlichen Mastputen*. (Dissertation). Hannover (2004).
45. Dalton HA, Wood BJ, Widowski TM, Guerin MT, Torrey S. Changes in leg health, skin, and plumage condition in domestic male turkeys of varying body weights. *Appl Anim Behav Sci.* (2016) 178:40–50. doi: 10.1016/j.applanim.2016.02.010
46. Bircher L, Hirt H, Oester H. Sitzstangen in der Mastputenhaltung. In: Kuratorium für Technik und Bauwesen in der Landwirtschaft, editor. *Aktuelle Arbeiten zur artgemäßen Tierhaltung: Vorträge anläßlich der 27. Internationalen Arbeitstagung Angewandte Ethologie bei Nutztieren der Deutschen Veterinärmedizinischen Gesellschaft, Fachgruppe Verhaltensforschung vom 22. - 25. 11. 1995 in Freiburg/Breisgau*. Darmstadt: KTBL (1995). p. 169–77.
47. Bellof G, Brandl M, Schmidt M, Carrasco S, Schade B. Einfluss unterschiedlicher Fütterungsintensität und Haltungsform auf die Mastleistung und den Schlachtkörperwert von langsam oder schnell wachsenden Genotypen in der ökologischen Putenmast. *Eur Poultry Sci.* (2014) 78. doi: 10.1399/eps.2014.31
48. Bellof G, Schmidt E. *Schlussbericht zum Forschungsprojekt Nr. 03OE451 – Einsatz ökologisch erzeugter Proteinträger in der Putenmast.* (2006). Available online at: http://orgprints.org/10902/1/10902-03OE451-fh-weihenstephan-schmidt-2006-putenmast.pdf
49. Habig C, Spindler B, Hartung J. *Gegenwärtige Management- und Haltungsbedingungen bei nicht schnabelgekürzten Puten in der ökologischen* Haltung: Abschlussbericht (2013). Available online at: http://www.ml.niedersachsen.de/download/91517/Abschlussbericht_Management-_und_Haltungsbedingungen_bei_unkupierten_Oekoputen.pdf
50. Hauck R, Hafez HM. Experimental infections with the protozoan parasite histomonas meleagridis: a review. *Parasitol Res.* (2013) 112:19–34. doi: 10.1007/s00436-012-3190-5
51. Buchwalder T, Huber-Eicher B. A brief report on aggressive interactions within and between groups of domestic turkeys (*Meleagris gallopavo*). *Appl Anim Behav Sci.* (2003) 84:75–80. doi: 10.1016/S0168-1591(03)00149-7
52. Savory CJ. Feather pecking and cannibalism. *World's Poultry Sci J.* (1995) 51:215–9. doi: 10.1079/WPS19950016
53. Bircher L, Schlup P. *Schlussbericht Teil 2 – Ethologische Indikatoren zur Beurteilung der Tiergerechtheit von Trutenmastsystemen*. Schlussbericht Bern. (1991).
54. Meyer H. Putenzucht im Überblick. In: Zentralverband der Deutschen Geflügelwirtschaft ev editor. *Geflügeljahrbuch 2019: Schwerpunkt: Biosicherheit und Hygiene*. Stuttgart: Eugen Ulmer Verlag (2018). p. 86–97.
55. Nestor KE, Bacon WL, Saif YM, Renner PA. The influence of genetic increases in shank width on body weight, walking ability, and reproduction of turkeys. *Poult Sci.* (1985) 64:2248–55. doi: 10.3382/ps.0642248
56. Tierschutzgesetz - in der Fassung der Bekanntmachung vom 18. Mai 2006 (BGBl. I S. 1206, 1313), das zuletzt durch Artikel 280 der Verordnung vom 19. Juni 2020 (BGBl. I S. 1328) geändert worden ist: TierSchG (2020).
57. International Society for Applied Ethology. *Guidelines for Ethical Treatment of Animals in Applied Animal Behaviour and Welfare Research*. (2017). Available online at: https://www.applied-ethology.org/res/EthicalGuidelinesISAErevised2017%20for%20council%20meeting.pdf

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.

frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 06 January 2021
doi: 10.3389/fvets.2020.613516

Walking on Tiptoes: Digital Pads Deserve Increased Attention When Scoring Footpad Dermatitis as an Animal Welfare Indicator in Turkeys

Jenny Stracke[1], Nina Volkmann[1], Franziska May[1], Stefanie Döhring[2], Nicole Kemper[1] and Birgit Spindler[1]*

[1] Institute for Animal Hygiene, Animal Welfare and Animal Behavior, University of Veterinary Medicine Hannover, Foundation, Hannover, Germany, [2] Faculty of Agricultural Sciences and Landscape Architecture, Osnabrück University of Applied Sciences, Osnabrück, Germany

Animal welfare is one of the most challenging issues in modern farm animal husbandry. Animal welfare indicators can be used to monitor welfare on farms or at slaughterhouses, with footpad dermatitis (FPD) being one of the most important indicators used in turkeys. Up to now, the severity of FPD has been measured by evaluating the size of altered lesions on the metatarsal pad of birds. However, such lesions are not only found on the metatarsal pads, but alterations can also occur on the digital pads of the animals, the latter is not included in the European standard scoring systems for turkeys so far. The aim of the present study was to give a detailed outline of alterations on the digital pads of turkeys and associate their occurrence to a standardly used five-point scoring system, which is based on alterations of the metatarsal pad only. Therefore, pictures of 500 feet of turkeys from 16 flocks at the end of the fattening phase were taken, using an automatic camera system. Based on these pictures, alterations on the digits were scored according to different parameters (lesions, swellings, and number of affected digits). Furthermore, detailed measurements were conducted using an imaging software. Results were compared with a standardly used five-point scoring system (standard FPD scoring system), based on the metatarsal pad as reference. Results provide no equivalence in occurrence and severity of alterations on the metatarsal pads compared to those found on the digits. Pathologic alterations on the digits were already present at standard FPD scoring level 0; no differentiation became obvious between the higher scoring levels 2–4. Strong correlations were found when comparing percentage of alterations of the standard FPD scoring system to those of a system including alterations on the digits and the metatarsal pad, using the total foot as a reference ($r_p = 0.9$, $p < 0.001$). This was the first study conducting a detailed analysis of alterations on the digits of turkeys. In conclusion, results of this study show that including the evaluation of alterations on digits could refine the present FPD scoring system, especially when using FPD as an animal welfare indicator.

Keywords: footpad dermatitis, FPD, turkey, digital pad, animal welfare, welfare indicator

OPEN ACCESS

Edited by:
Edward Narayan, The University of Queensland, Australia

Reviewed by:
Joanna Marchewka, Polish Academy of Sciences, Poland
Rosangela Poletto, Federal Institute of Rio Grande do Sul, Brazil

***Correspondence:**
Jenny Stracke
jenny.stracke@tiho-hannover.de

Specialty section:
This article was submitted to Animal Behavior and Welfare, a section of the journal Frontiers in Veterinary Science

Received: *02 October 2020*
Accepted: *04 December 2020*
Published: *06 January 2021*

Citation:
Stracke J, Volkmann N, May F, Döhring S, Kemper N and Spindler B (2021) Walking on Tiptoes: Digital Pads Deserve Increased Attention When Scoring Footpad Dermatitis as an Animal Welfare Indicator in Turkeys. Front. Vet. Sci. 7:613516. doi: 10.3389/fvets.2020.613516

INTRODUCTION

Animal welfare is not only of increasing concern in public but also gains importance in scientific and animal husbandry fields. In turkeys, FPD is one of the most frequently used welfare indicators providing information on animal health and well-being. This is due to the fact that FPD is closely associated with the husbandry system, with litter quality being the major determinant for this pathology (1–4). Furthermore, the manifestation of FPD is easy and quick to assess, implying its advantage to be utilized in various welfare assessment schemes, which are applied on-farm to measure animal welfare (5, 6).

FPD is described as a contact dermatitis of the plantar surface of birds' feet (7), which can show a wide range of characteristics. It occurs with different severity grades and can affect the surface but also subjacent structures (7, 8). Evaluation of severity is generally performed using scoring systems, categorizing the different incidents according to a subjective assessment of the size of the alteration. There are different scoring systems, which can be used in turkeys (1, 9, 10). However, a major breakthrough came in 2008, with Hocking et al. (11) proposing a standard scoring system to be used Europe-wide. This is especially important, as the monitoring of FPD at the slaughterhouse and on-farm is nowadays an accepted tool all over Europe and in the United States [see (9) for the United Kingdom and the United States; (11, 12) for Europe]. In order to ensure reliability (which is essential in a scientific context, but also plays a great role when focusing on international market competition), such a system, especially when practically used at the slaughterhouse, should be clearly defined, results should be repeatable between different classifiers and it should be quick and easy to use. The five-point scoring system proposed by Hocking et al. (11) fulfills all these criteria (11, 13, 14).

Nonetheless, this system also has its limitations, which become particularly important when using FPD not only as a benchmark system, but also as a welfare indicator, which is the current trend. In regard to animal welfare and to animal welfare legislation [(15), Article 7], unnecessary suffering should be avoided. Ulcerations can be considered as highly relevant in this context, as they are most likely to induce pain (1, 16, 17). Ulcerations are described as a loss of the epidermis, usually associated with an inflammatory reaction (18). According to a study by Stracke et al. (14), there is a link between the occurrence of ulcerations and the size of the lesion on the metatarsal pad, findings which are in agreement with Toppel et al. (13). Similar results were observed for broiler chickens (19, 20). However, using the standard scoring system of Hocking et al. (11), the study by Stracke et al. (14) also showed that no differences in severity of ulcerations were found between the higher scoring levels (scoring levels 2, 3, and 4). Furthermore, other incidents, such as re-epithelialized granulation tissue and chronic inflammation processes, were not to be found to be linked to the size of the lesion either. Stracke et al. (14) therefore raised the question as to whether the standardly used scoring system should be revised, at least with regard to the implementation of FPD as an animal welfare indicator. The present study takes a similar approach, albeit concentrating on another aspect, here in particular questioning the reference values used for evaluating FPD. Using the standard FPD scoring system, the metatarsal pad and its alterations are considered to represent incidence thereof. However, birds affected by FPD not only show alterations on the metatarsal pads. Alterations can also be found on the digital pads of the animals, as mentioned in various studies (1, 3, 7). Some existing studies state that the occurrence of FPD at the digits is most likely accompanied by higher scoring levels of measurements related to the metatarsal pad (21, 22). Up to now, there is no systematic study describing the development of lesions on the digital pads of turkeys. Consequently, the above-mentioned assumptions are not scientifically verified yet. Hocking et al. (11) also state that there are similar alterations at the toes compared to the ones found on the metatarsal pads and therefore, in order to obtain simplicity, they refrain from including the digits in their scoring systems. With regard to visual scoring systems, this might well be the case. However, with new forms of automatic technology for evaluating FPD, this approach might be worth reconsidering. Precision Livestock Farming (PLF) is rapidly developing in the poultry sector worldwide (23), thus offering opportunities to increase the efficiency and sustainability of farming and production and to improve animal health and welfare (24). Image analysis seems to be a promising approach to automatically evaluate FPD at the slaughterhouse, first systems of which have already been used for broilers (25, 26). In German slaughterhouses, a similar technique is employed for turkeys (13), with the classification of severity levels being based on the standard scoring system of Hocking et al. (11). In this latter study, the authors state that integrating alterations on the digits might improve the quality of FPD evaluation.

Therefore, in order to improve the standard FPD scoring system for usage in animal welfare measurements, the present study aimed to provide a detailed description of the lesions on digits. More specifically, the aim was to evaluate a potential linkage between lesions on the metatarsal and digital pads in turkey feet. According to current literature, the hypothesis is, that, with rising severity of FPD on the metatarsal pad, an increase in severity of alterations on digitals can be found.

METHODS

Turkey feet (B.U.T. 6, Aviagen Turkeys Ltd., Tattenhall, UK) were monitored at a German slaughterhouse at the end of the fattening phase. The outlined study was part of a larger project with the focus to validate the accuracy of automatic systems for the evaluation of FPD in turkeys at the slaughterhouse. The sample size used for the present study resulted from the original research question. Therefore, no a priori power analysis was conducted. In total, pictures of 16 flocks of male animals were taken using an automatic camera system (CLK GmbH; Turkey Check V1.0, Altenberge, Germany), which was a fixed part of the slaughter line used for continuously monitoring FPD. The camera system was installed at the end of the slaughter line, where feet are already separated from the body. It takes pictures of each foot passing the camera; one foot per pair of feet is used for further processing. Typically, the left foot per pair is used. However, if the left foot cannot be detected clearly, the camera switches to the right foot, respectively. The system, which is equipped with a software, is based on 2-D-RGB-image analysis and processing. Feet were detected using the contrast to a dark background. The software then checks the metatarsal pad for discoloration (darker areas). The metatarsal pad was defined as the difference between the foot and the digits, using the contrast between the brighter colors of the skin at the digits vs. the dark color of the background in the interspaces between the digits, leaving out the digits, putting a circle around the rest (see **Figure 1**).

FIGURE 1 | References for scoring. Total foot = RA1 is marked in red; Metatarsal pad = RA2 is marked in blue.

From the above mentioned flocks, 500 feet were picked in a pseudorandomized order, including 100 feet per scoring level of a five-point classification system [standard FPD scoring system, (11); see **Table 1**]. Feet were picked according to the scoring of the above mentioned camera system. Therefore, only the evaluated foot per pair was included in the sample, which could be either the right or the left one.

All pictures used for the analysis were verified manually regarding the performance of the automatic system. In a first step, a manual observer checked if the metatarsal pad and the alteration on the footpad had been correctly identified; in a second step, all footpads were scored manually, scoring levels (manual vs. automatic) having to be identical. Feet used for further analysis had to fulfill all of these selection criteria.

Scoring System FPD Metatarsal Pads

The definition of the different scoring levels can be found in **Table 1**. This scoring system is hereafter referred to as standard FPD scoring system.

Scoring System FPD Digital Pads

Only digital pads 2–4 were evaluated. This was due to practical reasons, as the suspension of feet in the slaughter line did not allow for the detection of the first digit. Each digital pad (DP) was scored separately, using a five-scale score to evaluate the altered lesion (**Table 2**) and a three-scale score was used, providing information on the severity grade of swelling (**Table 3**). Altered lesions were scored in relation to the different segments per digit. Furthermore, the number of affected digits was counted. The severity grade of swelling was evaluated using the corresponding digit of the respective (unaffected) second foot of each pair of feet as a reference.

TABLE 1 | Scoring system for the footpad dermatitis, based on alterations on the metatarsal pad adapted from Hocking et al. (13) (standard FPD scoring system) (pictures taken by Jenny Stracke).

Score 0	Score 1	Score 2	Score 3	Score 4
				
Intact foot	Small, punctual alterations, <10% of the footpad	Altered lesion covers ≤25% of the footpad	Altered lesion covers ≤50% of the footpad	Altered lesion covers more than 50% of the footpad

TABLE 2 | Scoring system for the digital pads: Altered lesion on the digital pad.

Score 0	Score 1	Score 2	Score 3	Score 4
				
Intact digit	Small, punctual alterations, on 1–2 segments and/or hematomas	Small, punctual alterations on >2 phalanxes or altered lesion in one phalanx which covers <50%	Altered lesion in one phalanx which covers more than 50% of the phalanx or several altered lesions in >one phalanx which cover <50%	Altered lesion in > one phalanx which covers ≥ than 50% of each phalanx or altered lesions covering an area larger than the phalanx

Digits (2–4) were scored separately, the digit used as example for the scoring system is highlighted (pictures taken by Jenny Stracke).

TABLE 3 | Scoring system for the digital pads: Grade of swelling on the digital pad.

Score 0	Score 1	Score 2
		
Intact digit	Slight swelling	Distinct swelling

Digits (2–4) were scored separately, affected digit is highlighted (pictures taken by Jenny Stracke).

Surface Measurements

A detailed survey of different parameters was conducted for 250 feet taken from the dataset. Here, the ImageJ software (Rasband, W.S., ImageJ, U. S. National Institutes of Health, Bethesda,MD, USA, https://imagej.nih.gov/ij/, 1997–2018) was used. Measurements were conducted to calculate the respective proportion of altered areas compared to different reference areas (RA). Therefore, the total foot (RA1), the metatarsal pad (RA2), the alteration on the footpad and the alterations on the digital pads (digital pad 2–4 in total) were tagged using the "freehand tool" in the software program (ImageJ, U. S. National Institutes of Health, Bethesda, MD, USA). RA1 included digital pads 2–4; the first digit was excluded as the suspension in the slaughter hooks did not allow a consistent presentation in the pictures of the automatic camera system. Furthermore, RA1 included the metatarsal pad, the remaining leg being excluded in the measurements (**Figure 1**). RA2 (metatarsal pad) was characterized by its kurtosis, the borders being specified at the start of the curvature (**Figure 1**). Alterations were defined as brownish discolorations, which could range from light to dark. The number of pixels on the surface area of both the reference and alteration was calculated to determine the respective proportion of the altered area compared to the respective reference (relative size of the lesion). All measurements were performed by one observer, observer reliability being ensured beforehand (see below for details). The percentage of the altered area in relation to the respective reference was calculated based on the following parameters:

- Alteration on the metatarsal pad in relation to the size of the metatarsal pad (standard)
- Alteration on the metatarsal pad in relation to the total foot (FP/RA1)
- Alteration on the digital pads in relation to the total foot (DP/RA1)
- Alteration on digital pads in relation to the metatarsal pad (DP/RA2)
- Alterations on metatarsal pad and digital pads in relation to the total foot (FULL)
- Size of the metatarsal pad in relation to the total foot (relative size of the metatarsal pad).

Observer Reliability

A separate dataset of digital pictures was used to test the observer reliability of the scoring system for detecting FPD (400 feet from two flocks, male animals, left and right feet). This sample was evaluated by two experienced observers (researcher/veterinarian). The applied scoring system can be found in **Table 1**.

Observer reliability for scoring the digital pads was calculated using a random sample (100 pictures) taken from the original subset. This sample was evaluated by two experienced observers (researcher/veterinarian). The scoring system can be found in **Tables 2**, **3**.

Observer reliability for the surface measurements was calculated using another dataset of 100 random pictures taken from the original dataset. Both observers (researcher) were experienced in using the program.

Statistical Analysis

For the statistical analysis, the SAS software (V.9.4, Statistical Analysis Institute, Cary, NC, USA) was used.

Observer reliability was calculated using the Krippendorff's alpha "macro" developed by Hayes and Krippendorff (27). The Krippendorff's alpha is a reliability coefficient, which in contrast to other reliability coefficients (e.g., the prevalence-adjusted and bias-adjusted kappa) does not only includes perfect agreements, but also takes into account the degree of discrepancies. This means that, if the given score levels differ only slightly (e.g., by one scoring level), the result would turn out better than if the score level difference is more pronounced (e.g., more than one scoring level) (28). The respective data type (ordinal data for scoring data; metric data for the measurements using ImageJ) was taken into consideration. Each data set was calculated separately. Observer reliability was evaluated using the classification proposed by Landis and Koch (29) (<0.00 = poor; 0.00–0.20 = slight; 0.21–0.40 = fair; 0.41–0.60 = moderate; 0.61–0.8 = substantial; 0.81–1.00 = almost perfect).

Scoring of the digital pads was analyzed by descriptive analysis using the FREQ procedure in SAS. Furthermore, principal component analysis (PCA) was conducted to condense the different parameters to one "digital score" in order to allow a comparison with the standard FPD scoring system. The PCA is a tool in multivariate statistics used for exploratory data analysis. It is commonly used for dimensionality reduction, by projecting data points to a few principal components and therefore, for obtaining lower-dimensional data while preserving as much of the data's variation as possible. The PCA is normally based on a matrix of Pearson's correlation of the original data, assuming that the variables are continuous. As our model included ordinal variables as well, a polychoric matrix was calculated first to serve as basis for the PCA. Here, the CORR procedure was applied using the polychoric option. The PCA was performed with the FACTOR procedure with the following parameter settings: method=PRINCIPAL, priors=SMC, rotation=VARIMAX. The number of extracted factors was specified using the mineigen statement (minimum eigenvalue) which was set to 1, therefore retaining components with an eigenvalue of 1 or >1. Corresponding PC scores for each foot were finally calculated with the SCORE procedure.

To analyze the relation between scoring of the digital pads with the standard FPD scoring system, a correlation analysis was performed using the CORR procedure, calculating the Spearman correlation coefficient between the standard FPD scoring system and the best factor resulting from the PCA. Furthermore, a generalized linear mixed model (GLIMMIX procedure) was calculated for this factor, including flock as a random effect. The standard FPD scoring system and the interaction between flock and the standard FPD scoring system were included as fixed effects, pairwise comparisons being conducted using Tukey-Kramer tests.

TABLE 4 | Observer reliability.

Parameter	Krippendorff's alpha
Standard scoring system (n = 400)	
Digital pictures of standard FPD scoring system	0.70
Scoring of the digital pads (n = 100)	
Altered lesions digital pad 2	0.82
Altered lesions digital pad 3	0.85
Altered lesions digital pad 4	0.84
Grade of swelling digital pad 2	0.61
Grade of swelling digital pad 3	0.80
Grade of swelling digital pad 4	0.67
Number of affected digits	0.83
Surface measurement	
Number of pixels on the total foot	0.97
Number of pixels on the metatarsal pad	0.84
Number of pixels of alterations on the metatarsal pad	0.83
Number of pixels of alterations on the digital pads (total)	0.69

To analyze differences between different digits regarding altered lesions and swellings, the Friedman test was calculated for each measurement separately using the ANOVA procedure (class variable: digital pads 2–4) in conjunction with the RANK procedure done by the blocking variable (foot).

The surface measurements were analyzed on a descriptive basis using the MEANS procedure in SAS. Furthermore, correlations were calculated using the CORR procedure, calculating the Pearson correlation coefficient between all parameters described above.

A generalized linear mixed model (GLIMMIX procedure) was calculated for the above mentioned parameters to analyze differences between the scoring levels of the standard scoring system, including the flock as random effect. Pairwise comparisons were conducted using Tukey-Kramer tests.

RESULTS

Observer reliability resulted in "moderate"—"almost perfect" results for all observations. Values of the Krippendorff's alpha coefficients for the different measurements can be found in **Table 4**.

Scoring FPD of the Digital Pads

In total, 17.8% of all feet (n = 500) were found to have intact digitals. In 27.2% of the feet, alterations and/or swelling became obvious on one digit, 28.4% of the feet showed alterations and/or swelling on two digits and in 26.6% of the feet, three digits were affected. **Figure 2** presents the results found for the number of affected digital pads per scoring level of the standard scoring

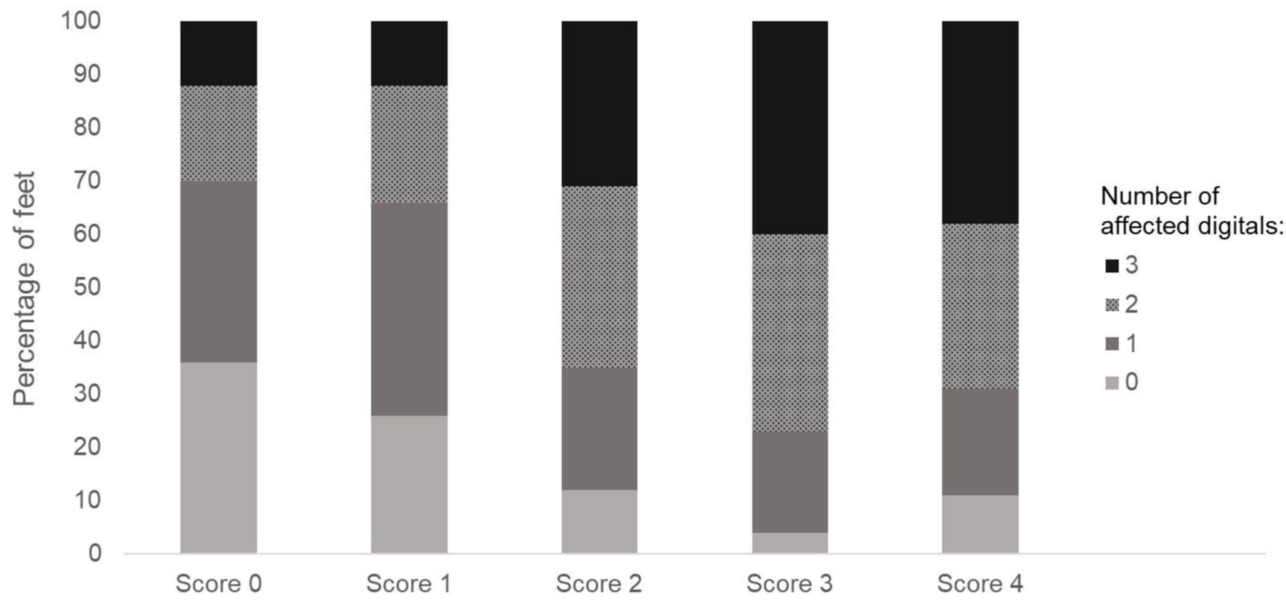


FIGURE 2 | Number of affected digits for the different scoring levels of the standard FPD scoring system of the metatarsal pad (score 0 = intact foot; score 1 = alterations ≤10%; score 2 = alterations ≤25%; score 3 = alterations ≤50% and score 4 = alterations >50%). Data are presented as percentage, *n* = 500 feet.

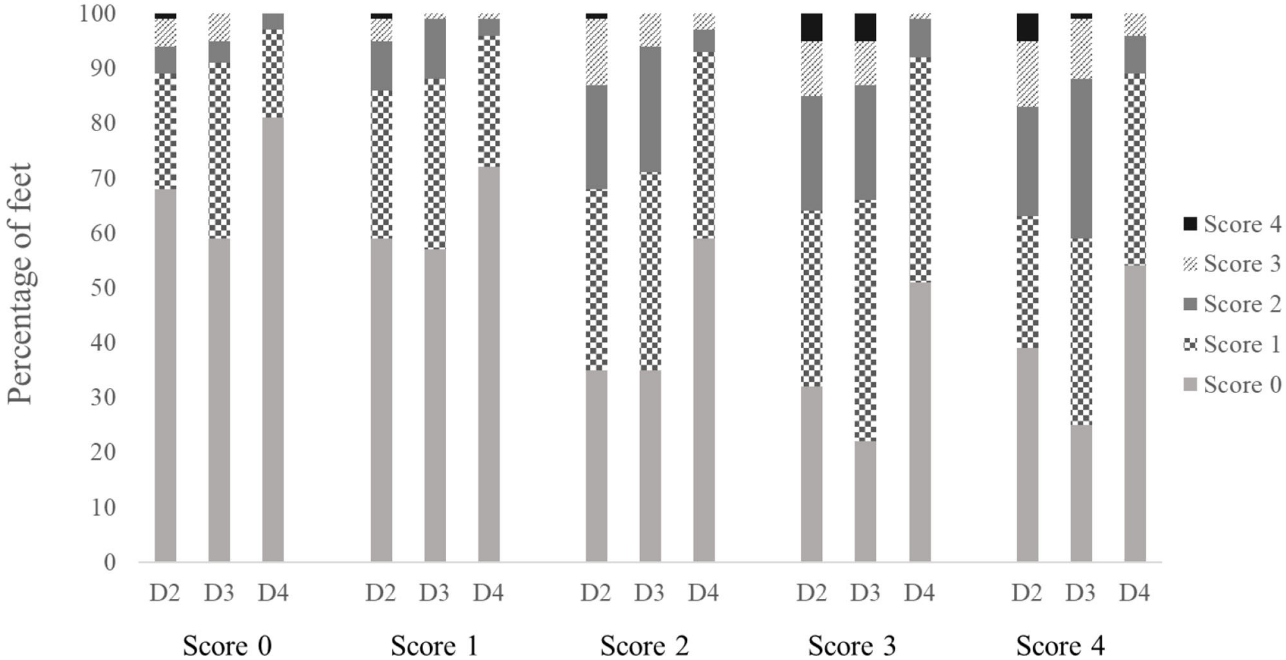


FIGURE 3 | Altered lesions on digital pads for the different scoring levels of the standard FPD scoring system of the metatarsal pad (score 0 = intact foot; score 1 = alterations ≤10%; score 2 = alterations ≤25%; score 3 = alterations ≤50% and score 4 = alterations >50%). Digits were scored separately (D2–D4). Data are presented as percentage, *n* = 500 feet.

scheme. Feet with the standard scoring level 0 ($n = 100$) were found to have one affected digital pad in 34.0% of the cases, 18.0% were found to have incidents on two digits, and in 12.0% of the cases, all three digits were affected. Scoring level 1 of the standard FPD scoring system was found to include 26.0% intact digits, in 40.0% of the cases, one digit was affected, whereas 22.0% of the cases showed alterations on two digits and in 12.0 % of the cases, alterations on three digital pads. In scoring level 2, the percentage of intact digital pads decreased to 12.0%, 23.0% of the feet were found to show alterations on one, 34.0% thereof on two and 31.0% on three digits. Scoring level 3 revealed 4.0% of the feet to have intact digital pads, with 19.0% of the feet showing alterations on one, 37.0% thereof alterations on two and 40.0% alterations on three digital pads. Results for scoring level 4 were similar, with 11.0% of the feet having no affected digit, 20.0% thereof one, 31.0% two and 38.0% three affected digital pads.

Figure 3 shows the results for the altered lesions found on each digit (D2–D4) in the different scoring levels of the standard FPD scoring system of the metatarsal pad. If affected, most lesions were allocated the severity grade 1 of the digital pad scoring system (standard FPD score: 0: 16–32%; 1: 24–31%; 2: 33–36%; 3: 32–44%; 4: 24–35%). Generally, severity grades of the digital score increased with an increasing severity in the standard FPD scoring system. However, differences between the standard FPD scoring levels 0 and 1 and between levels 2, 3 and 4 were only marginal (digital scoring levels for the standard score 0: 3–11%; 1: 4–14%; 2: 7–32%; 3: 8–36%; 4: 11–41%).

The results of the Friedman test revealed a significant difference between the digits [$F_{(2,1,461)} = 65.55$; $p < 0.001$], with digit 4 being affected least.

This could also be found for the grade of swelling [$F_{(2,1,461)} = 184.89$; $p < 0.001$]; here, digit 2 was affected most. Regarding the process of scoring, results revealed differences in the observer reliability, with the best values found for digit 3, whereas scoring for digit 2 and 4 only achieved moderate results.

Figure 4 presents the results for the grade of swelling found for each digital pad. Here again, severity grades of the digital score increased with an increasing severity in the standard FPD scoring system.

The PCA resulted in one factor with an eigenvalue >1 (2.8). Using this factor per foot (hereafter denoted as digital score) and the respective scoring level of the standard FPD scoring system for the correlation analysis revealed a moderate and significant positive correlation ($r_s = 0.41$; $p < 0.001$). A significant difference between scoring levels of the standard FPD scoring system could be found for the digital score [$F_{(4,448)} = 6.4$; $p < 0.001$], whereas pairwise comparisons revealed higher levels in scoring levels 2–4 compared to scoring level 1, whereas no significant differences could be found for the remaining combinations (**Figure 5**). Furthermore, a significant effect could be found for the interaction of the standard FPD scoring system and the flock [$F_{(35,448)} = 5.3$; $p < 0.001$]. Pairwise comparisons resulted in differences between particular flocks for the standard FPD scoring levels 2–4 (all $t < 8.0$; all $p < 0.05$), whereas no differences were found for scoring levels 0 and 1.

Surface Measurements

The results of the surface measurements can be found in **Figure 6**. For all parameters, a significant difference between scoring levels of the standard FPD scoring system could be found [standard system: $F_{(4,221)} = 493.8$; $p < 0.001$, FP/RA1: $F_{(4,229)} = 190.7$; $p < 0.001$), DP/RA2: $F_{(4,222)} = 4.3$; $p < 0.01$, DP/RA1: $F_{(4,229)} = 6.6$; $p < 0.001$, FULL: $F_{(4,229)} = 211.9$; $p < 0.001$]. Pairwise comparisons revealed significant differences between all scoring levels for the standard FPD scoring system (all $p < 0.05$), except for the scoring level 0 vs. scoring level

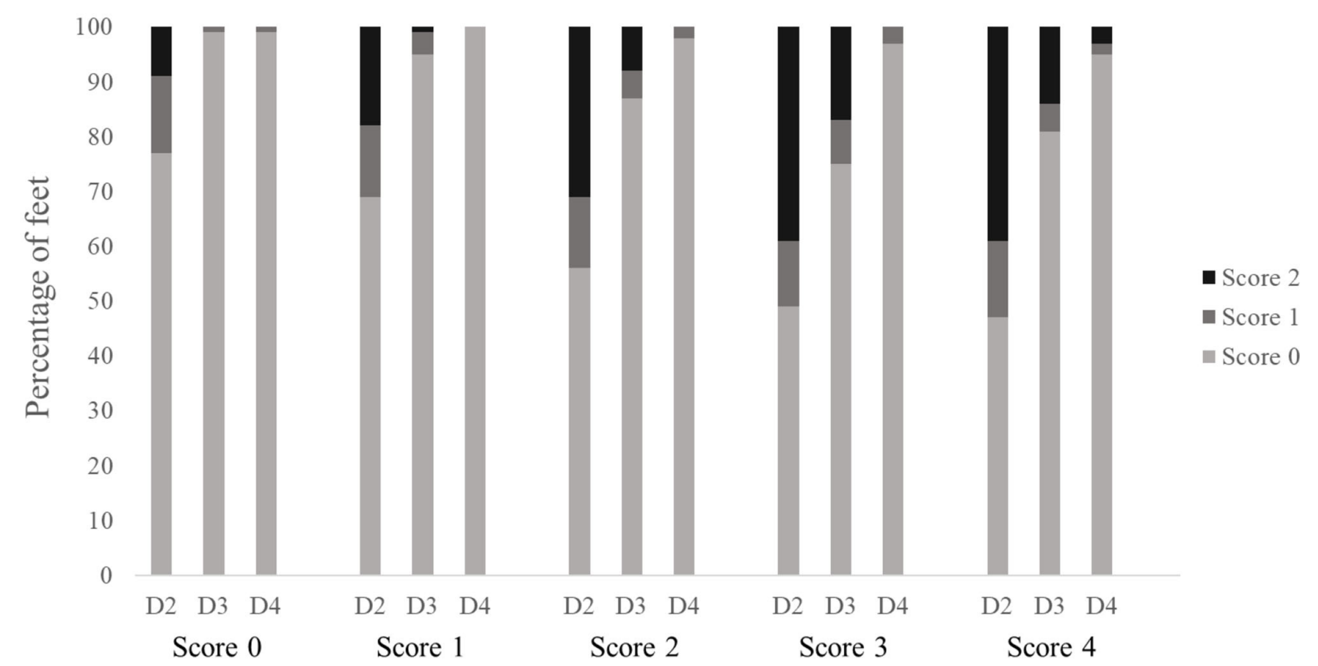


FIGURE 4 | Grade of swelling on digital pads for the different scoring levels of the standard FPD scoring system of the metatarsal pad (score 0 = intact foot; score 1 = alterations ≤10%; score 2 = alterations ≤25%; score 3 = alterations ≤50% and score 4 = alterations >50%). Digits were scored separately (D2–D4). Data are presented as percentage, $n = 500$ feet.

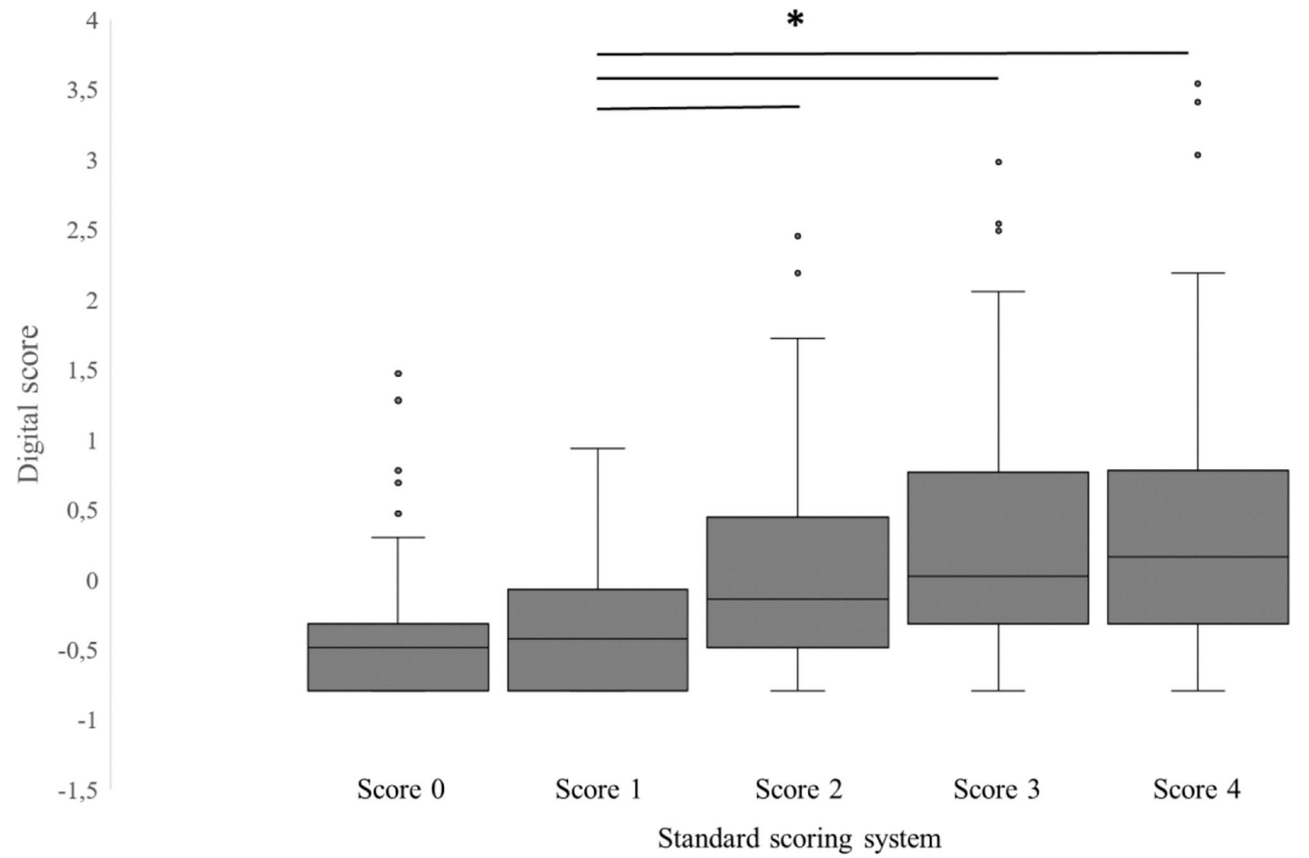


FIGURE 5 | Boxplots of the digital score [factor revealed from principal component analysis (PCA)] vs. the standard FPD scoring system of the metatarsal pad (score 0 = intact foot; score 1 = alterations ≤10%; score 2 = alterations ≤25%; score 3 = alterations ≤50% and score 4 = alterations >50%). Data are presented as boxplots (data range, median, lower quartile, and upper quartile; outliers are included in the graph as dots), *indicates a $p < 0.05$, $n = 500$ feet.

1. Here, a tendency could be found ($p = 0.056$). In FP/RA1, no significant differences between scoring levels 0 and 1 ($p = 0.53$) in contrast to all other combinations (all $p < 0.05$) were found. In DP/RA2, significant differences were found for the pairing of scoring levels 4 and 2 and scoring levels 4 and 1 only (all $p < 0.5$). For DP/RA1, significant differences were restricted to the differentiation between scoring level 4 and respective scoring levels (all $p < 0.05$). The FULL scoring system found no significant differences between scoring levels 0 and 1 ($p = 0.99$) in contrast to all other combinations (all $p < 0.05$).

The correlation coefficients between all parameters and the *p*-values of the correlation can be found in **Table 5**. A strong positive correlation was found between the standard FPD scoring system and alterations measured on the metatarsal pad in relation to RA1 ($p < 0.001$; $r_p 0.9$) and between the standard system and the FULL scoring system ($p < 0.001$; $r_p = 0.9$). The positive correlation between the standard FPD scoring system and measurements on the digital pads were only weak ($r_p = 0.2$ and 0.3; both $p < 0.001$).

The scoring of the digital pad in relation to the FULL scoring system resulted in a moderate and significant positive correlation ($p < 0.001$; $r_p = 0.5$).

The results for the relative size of the metatarsal pad revealed a significant difference between the scoring level of the standard FPD scoring system [$F_{(4,222)} = 59.6$; $p < 0.001$] (see **Figure 7**). The size increased significantly between severity grades (all $p < 0.05$), except for the comparison of scoring levels 1 and 2 ($p = 0.09$) and 2 and 3 ($p = 0.08$) where a tendency could be found.

DISCUSSION

The aim of this study was to evaluate alterations on the digital pads of turkeys and to compare the severity to those of alterations assessed on the metatarsal pads, evaluated by a standardized scoring system for FPD (11).

Even if included in some of the studies evaluating FPD in turkeys (1) and in assessment schemes in Germany (6), the digital pads are currently not included in the evaluation when using standard scoring systems for practical applications (9, 11). In contrast to turkeys, scoring systems used in broilers partially include lesions on the digital pads (20) though only in the highest severity grades. According to the literature (11, 21, 22), the assumption was that with rising severity grades of altered lesions

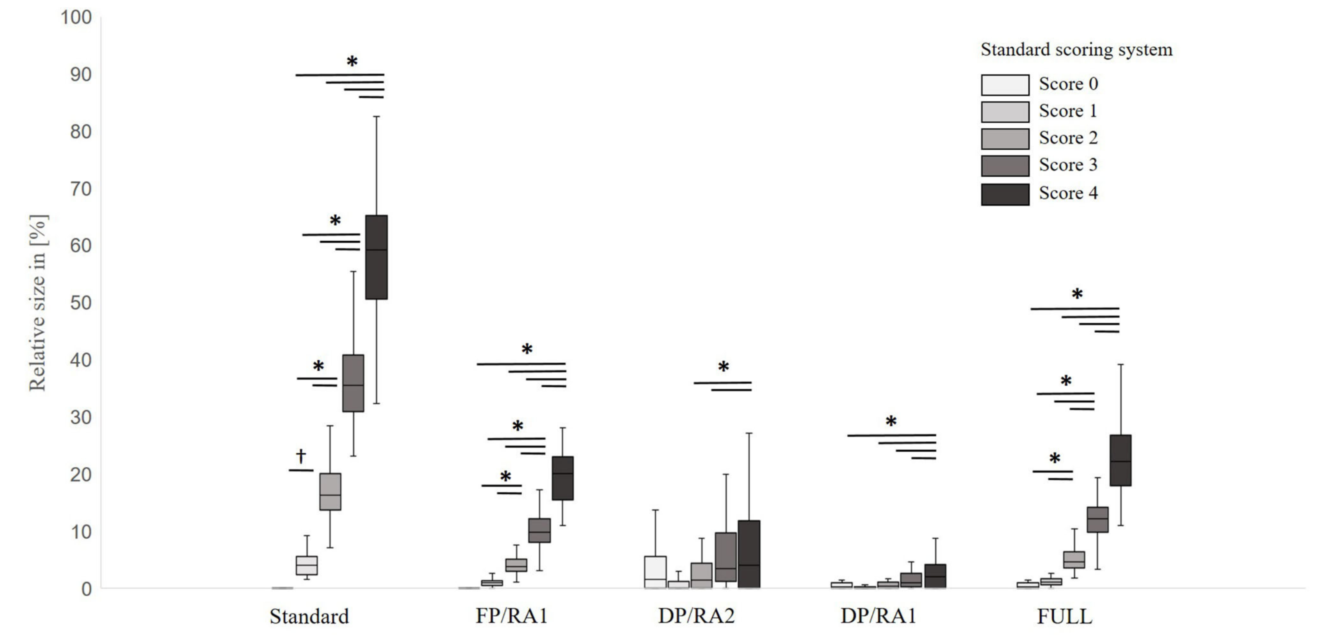


FIGURE 6 | Boxplots of the relative size of alterations measured for the different parameters by scoring level of the standard FPD scoring system. Data are presented as boxplots (data range, median, lower quartile and upper quartile; outliers are not included in the graph); $n = 250$, (Standard, alteration on the metatarsal pad in relation to the size of the metatarsal pad; FP/RA1, alteration on the metatarsal pad in relation to the total foot; DP/RA2, alteration on digital pads in relation to the metatarsal pad; DP/RA1, alteration on the digital pads in relation to the total foot; FULL, alterations on metatarsal pad and digital pads in relation to the total foot), *indicates a $p < 0.05$, †indicates a $p < 0.1$.

TABLE 5 | Correlation analysis between parameters of the surface measurement.

Parameters	r_p	*p*-value
Standard with		
FP/RA1	0.93	<0.001
DP/RA1	0.31	<0.001
DP/RA2	0.23	<0.001
FULL	0.91	<0.001
FP/RA1 with		
DP/RA1	0.17	<0.01
DP/RA2	0.09	*P*=0.17
FULL	0.92	<0.001
DP/RA1 with		
DP/RA2	0.84	<0.001
FULL	0.53	<0.001

The Pearson correlation coefficient (r_p) is presented. Standard, alteration on the metatarsal pad in relation to the size of the metatarsal pad; FP/RA1, alteration on the metatarsal pad in relation to the total foot; DP/RA1, alteration on the digital pads in relation to the total foot; DP/RA2, alteration on digital pads in relation to the metatarsal pad; FULL, alterations on metatarsal pad and digital pads in relation to the total foot.

on the metatarsal pad, alterations on digital pads would increase alike. If this would have been the case, the metatarsal pad could be used as a representative for the whole foot.

In contrast to the current opinion, the results of this study could not confirm a homogeneous representation of the severity of FPD when considering metatarsal pads and digitals separately. Even if the study generally found higher severity scores for lesions, swellings, the number of affected digits and the combination of all those parameters for the higher scoring levels of the standard FPD scoring system, no clear distinction became obvious between scoring levels 0 and 1 or between scoring levels 2–4. These findings were also confirmed by the surface measurements. When measuring the alterations on the digital pads, no differences in size could be found between standard scoring levels 0 and 1; numerical data even found higher values for scoring level 0 compared to scoring level 1. Furthermore, no differences in the size of the alteration were found between standard scoring levels 2 and 3. Additionally, correlations between measurements based on the standard system and measurements on the digital pads were only weak—also indicating a rather asynchronous occurrence of alterations on digital pads and the metatarsal pad. Especially alterations on digital pads of feet, which were scored as intact, according to the standard FPD scoring system, is relevant with regards to animal welfare. This effect could be due to missing parameters in the standard system. Stracke et al. (14) were able to demonstrate that other parameters like perivascular pododermatitis and re-epithelialized granulation tissue occur in feet scored as potentially intact, too. The occurrence of alterations on the digital pads therefore could be an effect of previous (old) metatarsal pad injuries. However, we are unable to ascertain which part of the foot was affected first. Further studies are needed to clarify the roots of this problem.

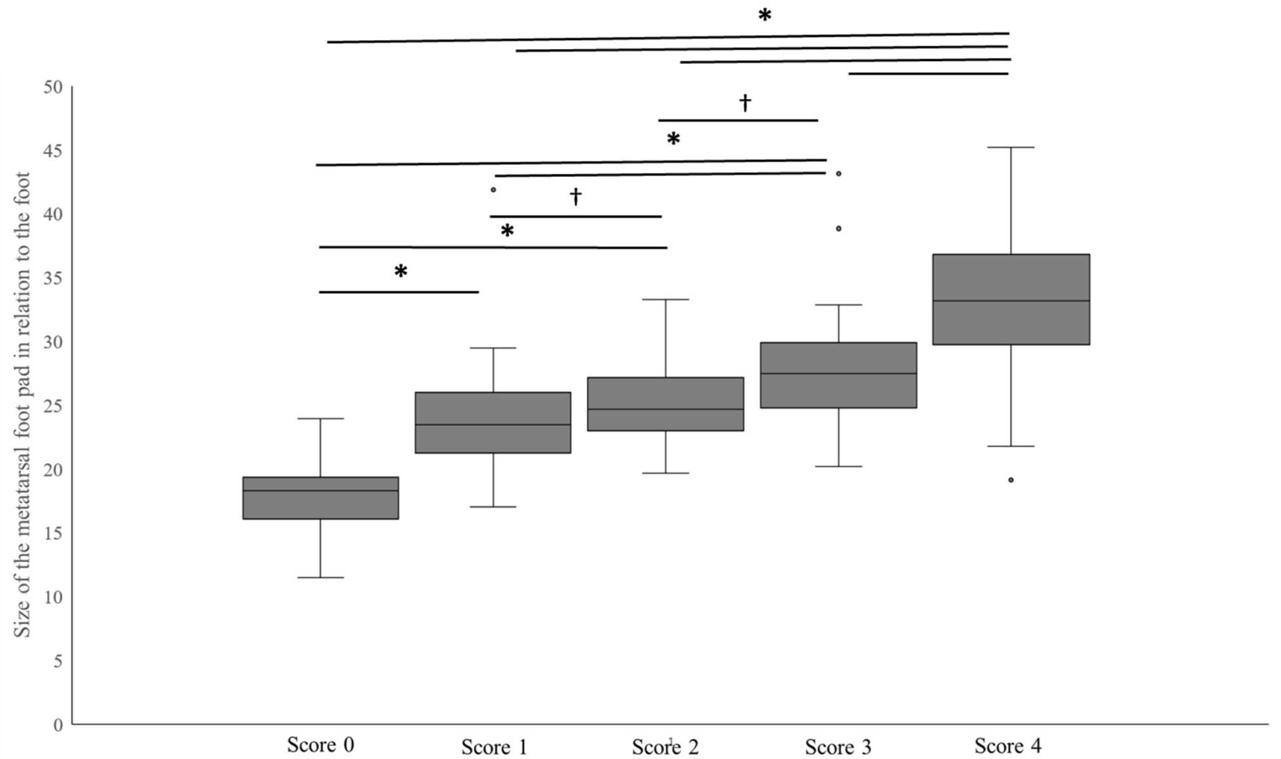


FIGURE 7 | Size of the metatarsal pad (independent of alterations) in relation to the size of the total foot for the scoring levels of the standard FPD scoring system. Data are presented as boxplots (data range, median, lower quartile, and upper quartile), *indicates a $p < 0.05$, †indicates a $p < 0.1$, $n = 242$ feet.

The most important reason for FPD in turkeys is the litter quality, with wet litter facilitating its occurrence (1–4). Feet evaluated in this study were picked randomly from a subset of 16 flocks; information on husbandry and litter quality of these flocks was not evaluated. Furthermore, standard FPD scores could not be streamlined for the different flocks, as not all scoring levels were present in all flocks. Therefore, it cannot be excluded that some effects found in the present study were due to varying husbandry conditions or restricted to specific flocks. The interaction effect between the flock and the standard FPD scoring system on the digital score, which could be found in this study, indicates no differences between flocks for standard scoring levels 0 and 1, whereas in standard scoring levels 2–4, differences between single flocks were present. However, more information on the background of flocks would be necessary to give a comprehensive picture in this case and to examine the causes of alterations on the digitals. This study only presents one first analysis of alterations on digitals and their linkage to a standard scoring system measuring the alterations on the metatarsal pad. Independent of the origin, the heterogeneous occurrence of alterations on digitals and metatarsal pads should be one hint, that scoring the metatarsal pads solely, might be insufficient to assess animal welfare.

Apart from that, while scoring the digital pads in this study, there was the subjective impression of finding a higher incidence of dirt under the nails of the animals showing higher severity grades in FPD scoring on the digital pads. However, we did not systematically evaluate this effect, and were unable to provide any information on correlations between dirtiness of the feet and FPD on the digital pads.

Both, metatarsal pad and digital pads are equipped with special fat structures (*Corpora adiposa plantaria superficialia et profunda*) (30, 31), which serve as mechanical protection from external pressure. These fat structures are more pronounced on the metatarsal pad; consequently, on the digital pads, there is less protection to the bones and underlying structures. It might therefore be plausible that birds start to relieve the digital pads by putting weight on the metatarsal pads instead when the digits are affected. This could be one explanation for the inconsistent development of alterations between the digital and metatarsal pads. However, this is highly speculative as there are no existing studies, neither evaluating the gait patterns of turkeys due to FPD in detail, nor the pressure load on specific body parts. There are results in laying hens providing evidence that there is a genetic influence on pressure load, which might be due to different weights of the animals (32), which might be evident in turkeys also (33, 34). That the pressure load might play a role in the occurrence of FPD on the digits is substantiated by the differences found between the digital pads concerning the alterations (digit 4 affected the least) and the grade of swelling (digit 4 affected the least, digit 2 affected the most). Both indicate an imbalanced pressure regarding the total foot. One explanation for this effect might be the leg position of the animals. Turkeys tend to be slightly bow-legged, leg problems like varus or valgus deviations can occur with a high prevalence, as a Danish study showed (35). Varus deviations were found to be correlated to weight in broiler chickens (36, 37), similar results being found for turkeys, too (38). We did not evaluate the weight of the animals in our study, but further studies might be beneficial in providing evidence on the development of FPD, including different gait patterns (e.g., by measuring the pressure load on specific body parts) and influences on other health parameters like for example weight gain (33, 34).

Generally, FPD is assumed to be painful (17, 39–41). However, relating the results of these previous studies to FPD by mainly using the gait score as parameter is not easy—other pathologies inducing an impaired gait, such as femoral head necrosis (42, 43) or osteomyelitis (44), being common in fast-growing poultry, too. Hocking et al. (45) could not prove higher FPD scores to be linked to painfulness in turkeys in a pharmacological study. In the context of pain, the grade of swelling might be a valuable parameter, as swelling is associated with inflammatory processes. Inflammation then again can be assumed to be linked to pain (46). Nevertheless, evaluating swelling from digital pictures as done in the presented study has to be interpreted with caution. As the results from the observer reliability confirm, the evaluation of the occurrence of swelling and the differentiation between slight and distinct was subjective. Defining the normal size and specified deviations from the norm proved to be extremely difficult due to differing size of feet and digits. In this case, other methods might be more feasible, like the manual palpation of fluctuation or thermal measurements, even if those techniques cannot be applied at the slaughterhouse. Apart from the grade of swelling, the present study found lesions in 82.2% of the analyzed feet. The study by Stracke et al. (14) was able to show that small lesions on the metatarsal pad can already be characterized by ulcerations; this is most likely to be the case for the lesions found on the digital pads as well. In mammals, ulcerations are referred to as being painful [see (47) for a review in pigs]. Therefore, the occurrence of ulcerations (on both, metatarsal pad and digital pads) should be considered a welfare aspect in turkeys, too [i.e., the concept of the Five Freedoms (Freedom from pain, injury and disease (48))].

In order to improve the assessment scheme for FPD with regard to animal welfare, it might be beneficial to include the monitoring of alterations on the digital pads. The present study found a high correlation between the FULL scoring system (including the digital pads) and the standard FPD scoring system; using the FULL system might therefore be an adequate alternative to the sole usage of the metatarsal pad. One critical point would be the easy application of the scoring system in situations where scoring has to be conducted fast, as scoring the total foot seems to be more demanding than simply scoring percentages on the metatarsal pad. However, observer reliability of alterations on the digital pads in the present study were good up to nearly perfect, implicating a good reliability of the scoring system *per se*. As an alternative, evaluating FPD on the digital pads could be conducted as part of an extended standard FPD scoring system. The involvement of digital pads could be implemented as an additional binomial score (alterations: yes/no) for all severity classes. This could also be beneficial with regard to the upcoming automatic assessment of FPD using 2-D-RGB-image analysis at the slaughterhouses. Even if the inclusion of scoring on the digits should be easy to apply, a separate evaluation of the digital pads would ease the manual evaluation in case of a technical breakdown. Besides implementing the digits, using the total foot as reference could refine existing automatic assessment methods, especially regarding recent discussions on the correct definition of the size of the metatarsal pad (13). As the results in this study provide strong correlations between the standard FPD scoring system and FP/RA1 (using alterations on the metatarsal pad in relation to the total foot), this could be a good alternative. Furthermore, the present study found the size of the metatarsal pad to increase with an increase in severity of the standard scoring. Similar effects were found in a study by Klambeck et al. (49) examining FPD in ducks. This effect might be due to inflammatory processes. Either way, whatever the underlying causes might be, such effects can falsify the assessment of FPD, when the reference for the assessment is based on the size of the metatarsal pad only. Using the total foot as a reference would not prevent such negative effects of swelling occurring, but could minimize the error rate. Furthermore, keeping the rapid development of automatic systems in mind, it might be worth thinking of possibilities of integrating the grade of swelling as well to gain a comprehensive picture of the pathology.

CONCLUSION

To conclude, the present study found no equivalent occurrence of alterations on the digital pads compared to alterations on the metatarsal pad assessed by a standard five-point scoring system. Pathologic alterations on the digital pads were present at standard scoring levels 0 already; no differentiation became obvious between the higher standard scoring levels 2–4. Good correlations were found when comparing the standard FPD scoring system to a system including alterations on the digital pads. Therefore, the authors state that including the digits could improve the present system with regard to animal welfare.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

Ethical review and approval was not required for the animal study because no living animals were included in this study. This study presents results based on digital pictures of turkey feet, samples base on recordings which were conducted during the standard process at the slaughterhouse.

AUTHOR CONTRIBUTIONS

BS and NK were responsible for financial acquisition and project development. JS, BS, and NK contributed to the conception and design of the present study. BS and JS were the principal investigators for this project. JS, NV, FM, and SD participated in data collection and performed the data analysis. JS designed and performed the statistical analysis. All authors were involved in interpreting the results and in drafting the manuscript.

FUNDING

This study was conducted within the project Autowohl—Automatic detection of animal welfare indicators (FKZ: 2817903515). The project was supported by funds from the Federal Ministry of Food and Agriculture (BMEL) based on a decision of the Parliament of the Federal Republic of Germany via the Federal Office for Agriculture and Food (BLE) under the innovation support program. The publication was supported by Deutsche Forschungsgemeinschaft, a German research funding organization, and the University of Veterinary Medicine Hannover, Foundation, Germany within the funding program Open Access Publishing.

ACKNOWLEDGMENTS

The authors wish to thank everyone involved in this study, especially Walter Wagner, Dr. Ronald Günther and CLK for their help during the practical performance and Rebekka Melloh for editing the example images of footpads. We also like to thank the reviewers for their suggestions.

REFERENCES

1. Martland MF. Wet litter as a cause of plantar pododermatitis, leading to foot ulceration and lameness in fattening turkeys. *Avian Pathol.* (1984) 13:241–52. doi: 10.1080/03079458408418528
2. Mayne RK. A review of the aetiology and possible causative factors of footpad dermatitis in growing turkeys and broilers. *World Poultry Sci J.* (2005) 61:256–67. doi: 10.1079/WPS200458
3. Mayne RK, Else RW, Hocking PM. High litter moisture alone is sufficient to cause footpad dermatitis in growing turkeys. *Br Poultry Sci.* (2007) 48:538–45. doi: 10.1080/00071660701573045
4. El-Wahab AA, Beineke A, Beyerbach M, Visscher CF, Kamphues J. Effects of floor heating and litter quality on the development and severity of foot pad dermatitis in young turkeys. *Avian Dis.* (2011) 55:429–34. doi: 10.1637/9684-021011-Reg.1
5. Welfare Quality R. *Welfare Quality R Assessment Protocol for Poultry (Broilers, Laying Hens).* Lelystad, Netherlands: Welfare Quality R Consortium. (2009).
6. Knierim U, Andersson R, Keppler C, Petermann S, Rauch E, Spindler B, et al. *Tierschutzindikatoren: Leitfaden für die Praxis—Geflügel.* Darmstadt: Kuratorium für Technik und Bauwesen in der Landwirtschaft (KTBL). (2016).
7. Greene JA, McCracken RM, Evans RT. A contact dermatitis of broilers—clinical and pathological findings. *Avian Pathol.* (1985) 14:23–38. doi: 10.1080/03079458508436205
8. Martrenchar A, Boilletot E, Huonnic D, Pol F. Risk factors for foot-pad dermatitis in chicken and turkey broilers in France. *Prev Vet Med.* (2002) 52:213–26. doi: 10.1016/S0167-5877(01)00259-8
9. Clark S, Hansen G, McLean P, Bond P Jr, Wakeman W, Buda S. Pododermatitis in turkeys. *Avian Dis.* (2002) 46:1038–44. doi: 10.1637/0005-2086(2002)0461038:PIT2.0.CO;2
10. Mayne RK, Else RW, Hocking PM. High dietary concentrations of biotin did not prevent foot pad dermatitis in growing turkeys and external scores were poor indicators of histopathological lesions. *Br Poultry Sci.* (2007) 48:291–8. doi: 10.1080/00071660701370509
11. Hocking PM, Mayne RK, Else RW, French NA, Gatcliffe J. Standard European footpad dermatitis scoring system for use in turkey processing plants. *World Poultry Sci J.* (2008) 64:323–8. doi: 10.1017/S0043933908000068
12. Pritchard DG. The impact of legislation and assurance schemes on alternative systems for poultry. In: Sandilands V, Hocking PM, editors. *Alternative systems for poultry: health, welfare and productivity.* Oxfordshire; Cambridge: Cabi (2012). p. 23–53. doi: 10.1079/9781845938246.0023
13. Toppel K, Spindler B, Kaufmann F, Gauly M, Kemper N, Andersson R. Foot pad health as part of on-farm-monitoring in turkey flocks. *Front Vet Sci.* (2019) 6:25. doi: 10.3389/fvets.2019.00025
14. Stracke J, Klotz D, Wohlsein P, Döhring S, Volkmann N, Kemper N, et al. Scratch the surface: histopathology of footpad dermatitis in turkeys (*Meleagris gallopavo*). *Animal Welfare.* (2020) 29:419–32. doi: 10.7120/09627286.29.4.419
15. European Convention for the Protection. *European Convention for the Protection of Animals kept for Farming Purposes*. (1976). Avaialble online at: https://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/087 (accessed 30 April, 2020).
16. Haslam SM, Knowles TG, Brown SN, Wilkins LJ, Kestin SC, Warriss PD, et al. Factors affecting the prevalence of foot pad dermatitis, hock burn and breast burn in broiler chicken. *Br Poultry Sci.* (2007) 48:264–75. doi: 10.1080/00071660701371341
17. Weber Wyneken C, Sinclair A, Veldkamp T, Vinco LJ, Hocking PM. Footpad dermatitis and pain assessment in turkey poults using analgesia and objective gait analysis. *Br Poultry Sci.* (2015) 56:522–30. doi: 10.1080/00071668.2015.1077203
18. Wohlsein P, Baumgärtner W, Hewicker-Trautwein M. Haut. In: Baumgärtner W, Gruber AD, editors. *Spezielle Pathologie für die Tiermedizin.* Stuttgart: Enke Verlag (2015). p. 342–9. doi: 10.1055/b-003-125811
19. Heitmann S, Stracke J, Petersen H, Spindler B, Kemper N. First approach validating a scoring system for foot-pad dermatitis in broiler chickens developed for application in practice. *Prev Vet Med.* (2018) 154:163. doi: 10.1016/j.prevetmed.2018.03.013
20. Piller A, Bergmann S, Schwarzer A, Erhard M, Stracke J, Spindler B, et al. Validation of histological and visual scoring systems for foot pad dermatitis in broiler chicken. *Animal Welfare.* (2020) 29:185–96. doi: 10.7120/09627286.29.2.185
21. Bergmann S, Ziegler N, Bartels T, Hübel J, Schumacher C, Rauch E, et al. Prevalence and severity of foot pad alterations in German turkey poults during the early rearing phase. *Poultry Sci.* (2013) 92:1171–6. doi: 10.3382/ps.2012-02851
22. Hafez HM, Rudolf M, Haase S, Hauck R, Behr KP, Bergmann V, et al. Influence of stocking density and litter material on the incidence of pododermatitis of turkeys. In: *Proceedings of 3rd International Symposium on Turkey Production: Prospects on Future Developments.* Berlin (2005). p. 101109.
23. Rowe E, Dawkins MS, Gebhardt-Henrich SG. A systematic review of precision livestock farming in the poultry sector: is technology focussed on improving bird welfare? *Animals.* (2019) 9:614. doi: 10.3390/ani9090614
24. Berckmans D. Precision livestock farming technologies for welfare management in intensive livestock systems. *Rev Sci Tech.* (2014) 33:189–96. doi: 10.20506/rst.33.1.2273
25. Ben Sassi N, Averós X, Estevez I. Technology and poultry welfare. *Animals.* (2016) 6:62. doi: 10.3390/ani6100062
26. Vanderhasselt RF, Sprenger M, Duchateau L, Tuyttens FAM. Automated assessment of footpad dermatitis in broiler chickens at the slaughter-line: Evaluation and correspondence with human expert scores. *Poultry Sci.* (2013) 92:12–18. doi: 10.3382/ps.2012-02153

27. Hayes AF, Krippendorff K. Answering the call for a standard reliability measure for coding data. *Commun Methods Measure.* (2007) 1:77–89. doi: 10.1080/19312450709336664
28. Krippendorff K. *Computing Krippendorff's alpha-reliability*. (2011). Available online at: http://repository.upenn.edu/asc_papers/43 (accessed 20 November, 2019).
29. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* (1977) 33:159–74. doi: 10.2307/2529310
30. Spearman RIC, Hardy JA. *Form and Function in Birds. Integument. Bd. 3.* London: Academic Press Inc. Ltd. (1985). p. 1–14.
31. Vollmerhaus B, Sinowatz F. Abstammung des Nutzgeflügels. In: Nickel RA, Schummer A, Seiferle E, editors. *Lehrbuch der Anatomie der Haustiere, Band 5: Anatomie der Vögel. 2.* Berlin, Hamburg: Auflage, Verlag Paul Parey (1992). p. 1–12.
32. Pickel T, Schrader L, Scholz B. Pressure load on keel bone and foot pads in perching laying hens in relation to perch design. *Poultry Sci.* (2011) 90:715–24. doi: 10.3382/ps.2010-01025
33. Oviedo-Rondón EO, Lascelles BDX, Arellano C, Mente PL, Eusebio-Balcazar P, Grimes JL, et al. Gait parameters in four strains of turkeys and correlations with bone strength. *Poultry Sci.* (2017) 96:1989–2005. doi: 10.3382/ps/pew502
34. Kremer JA, Robison CI, Karcher DM. Growth dependent changes in pressure sensing walkway data for Turkeys. *Front Vet Sci.* (2018) 5:241. doi: 10.3389/fvets.2018.00241
35. Sanotra GS, Lund JD, Ersøll AK, Petersen JS, Vestergaard KS. Monitoring leg problems in broilers: a survey of commercial broiler production in Denmark. *World Poultry Sci J.* (2001) 57:55–69. doi: 10.1079/WPS20010006
36. Crespo R. Developmental, metabolic, and other non-infectious disorders. *Dis Poultry.* (2020) 1286–329. doi: 10.1002/9781119371199.ch30
37. Bihan Duval EL, Beaumont C, Colleau JJ. Estimation of the genetic correlations between twisted legs and growth or conformation traits in broiler chickens. *J Anim Breeding Genetics*. (1997) 114:239–59. doi: 10.1111/j.1439-0388.1997.tb00510.x
38. Kapell DNRG, Hocking PM, Glover PK, Kremer VD, Avendaño S. Genetic basis of leg health and its relationship with body weight in purebred turkey lines. *Poultry Sci.* (2017) 96:1553–62. doi: 10.3382/ps/pew479
39. Martland MF. Ulcerative dermatitis in broiler chickens: the effects of wet litter. *Avian Pathol.* (1985) 14:353–64. doi: 10.1080/03079458508436237
40. Hocking PM, Wu K. Traditional and commercial turkeys show similar susceptibility to foot pad dermatitis and behavioural evidence of pain. *Br Poultry Sci.* (2013) 54:281–8. doi: 10.1080/00071668.2013.781265
41. Sinclair A, Weber Wyneken C, Veldkamp T, Vinco LJ, Hocking PM. Behavioural assessment of pain in commercial turkeys (*Meleagris gallopavo*) with foot pad dermatitis. *Br Poultry Sci.* (2015) 56:511–21. doi: 10.1080/00071668.2015.1077204
42. Dinev I. Clinical and morphological investigations on the prevalence of lameness associated with femoral head necrosis in broilers. *Br Poultry Sci.* (2009) 50:284–90. doi: 10.1080/00071660902942783
43. Packialakshmi B, Rath NC, Huff WE, Huff GE. Poultry femoral head separation and necrosis: a review. *Avian Dis.* (2015) 59:349–54. doi: 10.1637/11082-040715-Review.1
44. Wyers M, Cherel Y, Plassiart G. Late clinical expression of lameness related to associated osteomyelitis and tibial dyschondroplasia in male breeding Turkeys. *Avian Dis.* (1991) 35:408–14. doi: 10.2307/1591199
45. Hocking PM, Harkness A, Veldkamp T, Vinco LJ. Do foot pad scores measure turkey welfare. In: *Proceedings of the 11th Turkey Science and Production Conference*. Chester (2017). p. 20–3.
46. Zhang JM, An J. Cytokines, inflammation and pain. *Int Anesthesiol Clin.* (2007) 45:27. doi: 10.1097/AIA.0b013e318034194e
47. Ison SH, Clutton RE, Di Giminiani P, Rutherford K. A review of pain assessment in pigs. *Front Vet Sci.* (2016) 3:108. doi: 10.3389/fvets.2016.00108
48. Farm Animal Welfare Council. *Farm Animal Welfare in Great Britain: Past, Present and Future*. (2009). Available online at: http://www.fawc.org.uk (accessed 30 April, 2020).
49. Klambeck L, Stracke J, Spindler B, Klotz D, Wohlsein P, Schoen HG, et al. First approach to validate a scoring system to assess footpad dermatitis in Pekin ducks. *Eur Poultry Sci.* 83:262. doi: 10.1399/eps.2019.262

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.

frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 06 January 2021
doi: 10.3389/fvets.2020.601304

Optimisation of Scores Generated by an Online Feline Health–Related Quality of Life (HRQL) Instrument to Assist the Veterinary User Interpret Its Results

Vinny Davies[1], *Jacqueline Reid*[2]* *and E. Marian Scott*[3]

[1] School of Computing Science, University of Glasgow, Glasgow, United Kingdom, [2] School of Veterinary Medicine, University of Glasgow, Glasgow, United Kingdom, [3] School of Mathematics & Statistics, University of Glasgow, Glasgow, United Kingdom

OPEN ACCESS

Edited by:
Alan G. McElligott,
City University of Hong Kong,
Hong Kong

Reviewed by:
Lauren Finka,
Nottingham Trent University,
United Kingdom
Janice Vaz,
Western Sydney University, Australia
Jenna Kiddie,
Dogs Trust, United Kingdom

***Correspondence:**
Jacqueline Reid
jacky.reid@newmetrica.com

Specialty section:
This article was submitted to
Animal Behavior and Welfare,
a section of the journal
Frontiers in Veterinary Science

Received: 31 August 2020
Accepted: 01 December 2020
Published: 06 January 2021

Citation:
Davies V, Reid J and Scott EM (2021)
Optimisation of Scores Generated by
an Online Feline Health–Related
Quality of Life (HRQL) Instrument to
Assist the Veterinary User Interpret Its
Results. Front. Vet. Sci. 7:601304.
doi: 10.3389/fvets.2020.601304

Using methodology previously described for the dog health-related quality of life (HRQL) tool (VetMetrica^{TM}), the aim was to optimize the scores profile of a comparable feline online HRQL instrument for monitoring HRQL in cats, to assist in its interpretation. Measuring HRQL helps quantify the impact of disease and its treatment on well-being, aids clinical decision making and provides information in clinical trials. In Study 1, using data collected from previous studies, scores generated for three domains of HRQL (Vitality, Comfort, Emotional Well-being) in healthy cats were normalized using standard statistical techniques of logit transformation and T-scores, such that the average healthy cat has a score of 50 in all three HRQL domains. Using normalized scores from healthy and sick cats, a threshold score of 44.8 was determined, above which 70% of healthy cats should score. Study 2 determined the Minimal Important Difference (MID) in normalized score that constituted a clinically significant improvement in each domain. Three methods were tested in order to determine the MID, with the final choice made based on statistical and clinical considerations. Thresholds of 5, 7.5, and 5 were chosen for the three HRQL domains representing Vitality, Comfort and Emotional Well-being, respectively. This study makes available a means of displaying HRQL scores from an online application in an easily interpretable manner and quantifies a clinically meaningful improvement in score. To illustrate the practical application of these developments, three case examples are presented. Example 1 illustrates the raw and normalized scores for a group of overweight cats enrolled in a Feline Weight Management Programme. Example 2 shows three groups of osteoarthritic cats, each with different severity of disease. The third is an elderly, un-well cat whose HRQL was recorded over time, specifically to facilitate end of life discussion between owner and veterinary clinician.

Keywords: health-related quality of life, cats, optimisation, health status, threshold, minimum important difference, interpretability, score normalization

INTRODUCTION

In humans HRQL measurement is an important research area, quantifying the impact of disease and its treatment on an individual's daily well-being, to aid clinical decision making, provide an outcome measure in clinical trials and contribute to healthcare policy (1). The development of HRQL instruments for companion animals is growing and their value is increasingly recognized (2). HRQL instruments can be disease specific or generic (3–8). Generic instruments measure quality of life (QOL) in healthy or sick animals and are the only option when co-morbidities are present, as is the case with older animals. Profile rather than single item HRQL measures generate scores in multiple domains of HRQL, allowing for comprehensive analysis of HRQL changes over time, in contrast to a single item score which only tells us if an animal is better or worse.

Previously we reported the development, validation and reliability of web-based generic HRQL profile measures for the dog (9, 10) and the cat (11). They consist of simple behavioral questions (questionnaire items) for the owner (22 for the dog and 20 for the cat) that are scored on a 0-6 scale (0 = could not be less and 6=could not be more) and completed online in around 5 min. A list of these items which are either positive—for example "active," or negative—for example "listless," is available in Noble et al. (11). The dog tool generates scores in four domains of HRQL -Energetic/Enthusiastic (E/E), Happy/Content (H/C), Active/Comfortable (A/C), Calm/Relaxed (C/R)) compared with 3 for the cat–Vitality, Comfort, Emotional Well-being (EWB). These domains were derived using a multivariate statistical analysis called factor analysis (FA). Factor analysis is a technique that is used to reduce a large number of variables into fewer numbers of factors, in this case domains of HRQL. Details of this process for the dog and cat tools can be found in Reid et al.'s and Noble et al. (9–11), but briefly, using a test data set, several factor solutions are explored, each of which consists of a different number of factors. The optimum factor solution is the one that accounts for the most variability in the data, so for the dog, the four factor solution was optimum accounting for 72% of the variance (10) and in the cat the optimum solution contained three factors, accounting for 72.3% of the variance (11).

Although interpretability is a key element of a useful measurement scale, to date there is no agreement as to how HRQL scores should be presented to ensure the ease with which a user can interpret them. Users should be able to understand what an individual score produced by an instrument means, for example by comparing the score to a healthy population. They also need to know when any variation in scores is meaningful, for example is the treatment really working or is the disease really worsening as time goes on? This is currently an important research focus in the medical field (12), where according to these authors "the choice of what constitutes an important difference in a HRQL score will influence judgements about the

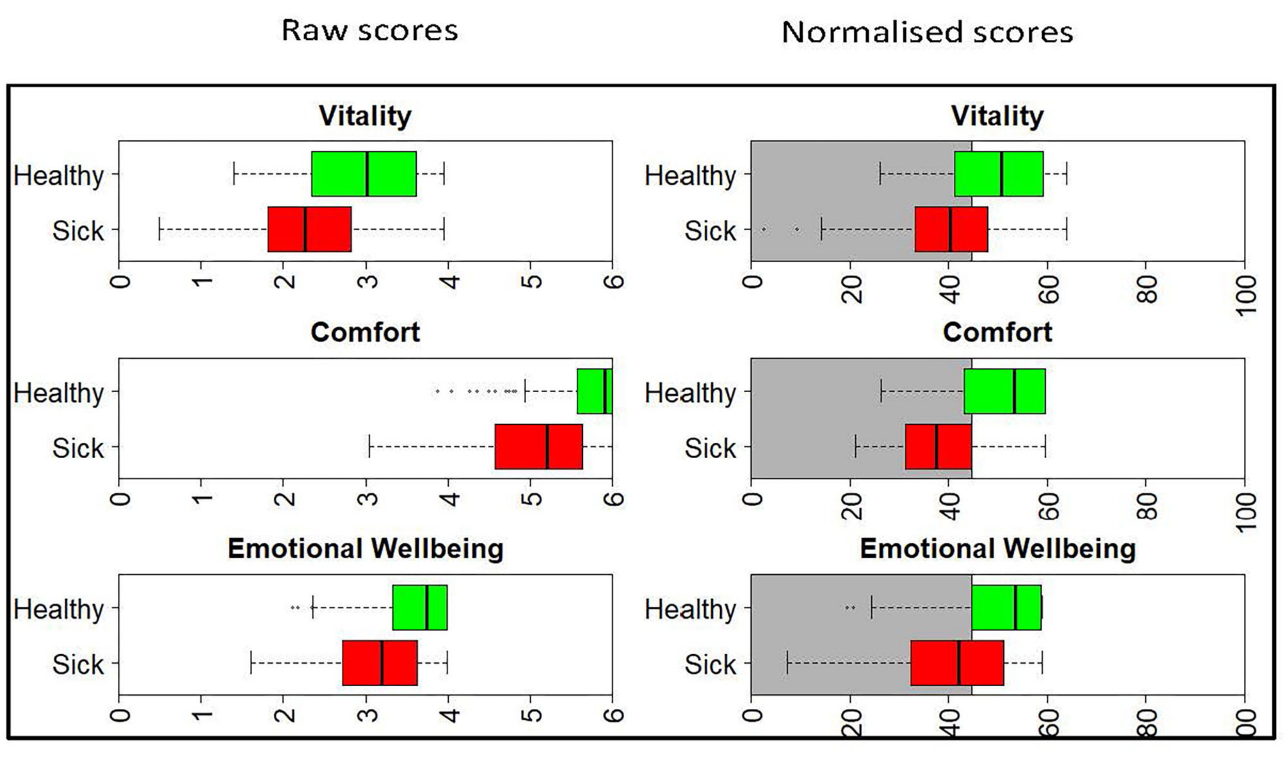


FIGURE 1 | Boxplots showing the raw and normalized HRQL domain scores for sick (red) and healthy (green) cats. The threshold for the normalized scores is set at 44.8 for all domains, with 70% of healthy cats estimated to score above this threshold. The shaded gray area represents the area below the 44.8 threshold score.

success of a health care intervention, the required sample size of clinical studies, and the design of these studies." Health—related quality of life instrument development is an expanding area in the veterinary arena and there is increasing appreciation of their value in measuring well-ness within a model of veterinary preventative healthcare and measuring the impact of chronic disease and its treatment (10). As a result, interpretability is equally as important in the veterinary as well as the medical field.

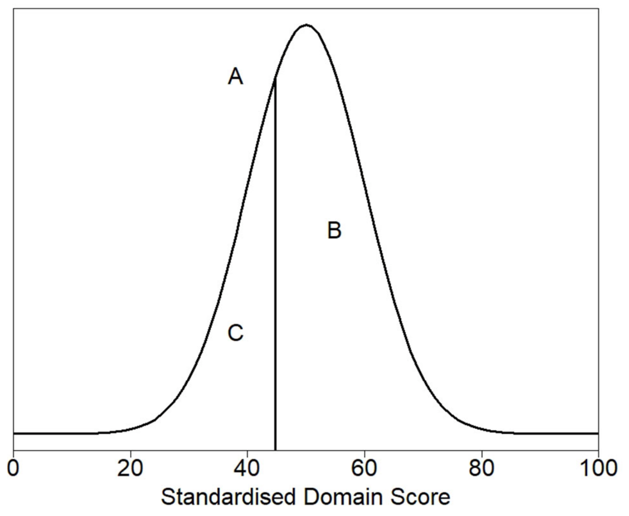


FIGURE 2 | Density plot of the theoretical distribution of the healthy cat population. **(A)** denotes the threshold for dividing healthy and sick cats, the 30th percentile (a HRQL score of 44.8). Above this threshold (to the right of the plot) are 70% of the healthy cat population predicted to be healthy, the area marked by **(B)**. Below the threshold (to the left of the plot) are 30% of the healthy population predicted to be sick, the area marked by **(C)**.

Currently, several techniques have been suggested to enhance interpretability of health measures for people including linking the scores to those of a specific population (norm-based scoring) such as a general population, populations with similar demographics or to a population with a specific disease (13). The scores profile of the dog tool was optimized to improve its interpretability by normalizing the scores to the age-related healthy dog population and deriving a threshold as a guide to the health of the dog (14). A detailed description of the rationale underlying the choice of these procedures can be found in Davies et al. (14). The significance of a change in score can be quantified through the calculation of a Minimum important difference (MID). This has been defined as "the smallest difference in score in the outcome of interest that informed patients or informed proxies perceive as important, either beneficial or harmful, and which would lead the patient or clinician to consider a change in the management" (15). The MID can be established using distributional and/or anchor-based techniques. A distribution-based approach relies on the statistical properties of the instrument and does not involve an external impression of change. Examples include effect size (16), normalized response mean (17) and the modified normalized response mean (18). On the other hand, anchor-based techniques use an external impression such as the patient's perception of a significant improvement or worsening of their condition to identify the change on the HRQL scale that corresponds to the MID (19). Such global measures of change are however strongly affected by the context in which they are used and subject to much variability making dependance on these problematic. Deyo and Centor (20) suggested that scales could be viewed as "diagnostic

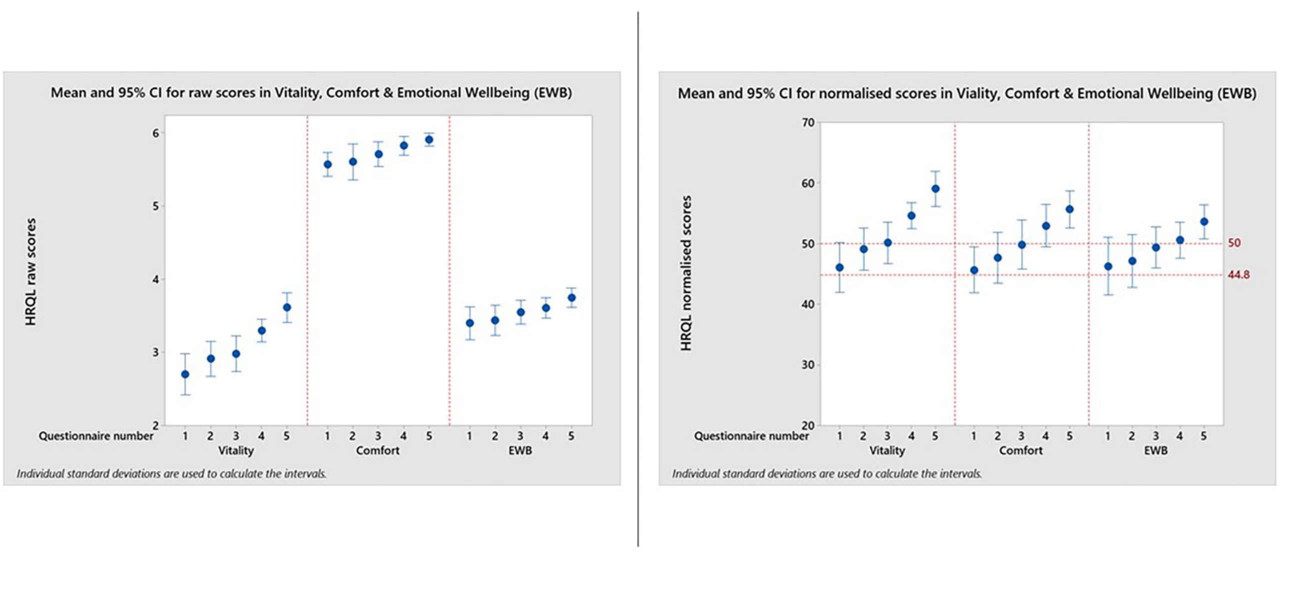


FIGURE 3 | Reproduced with kind permission of Dr. Barr Hadar, Ontario Veterinary College, University of Guelph. Means and 95% confidence intervals for HRQL scores in Vitality, Comfort, Emotional Well-being for a group of 21 overweight cats enrolled on a weight loss programme whose owners completed a baseline score before treatment and then four assessments at approximately monthly intervals. A, Raw domain scores; B, Normalized domain scores. A score of 50 = the average healthy cat score and 70% of healthy cats will score above the 44.8 threshold.

tests" for distinguishing improved patients from those that had not, with receiver operating characteristic (ROC) curves being utilized to describe a scale's ability to identify improvement. This more objective method to determining the MID using ROC was used in the dog study (14). Briefly, the characteristics of the test, namely sensitivity and specificity, which describe how well the test discriminates between two groups are calculated. Sensitivity describes how well a test identifies those with a particular disease (true positive) and specificity describes how well it correctly identifies those without that disease (true negative). The ROC curve plots sensitivity against 1—specificity and a cut-off (threshold value) is chosen above which cases are classified as positive while cases with scores below that cut-off are classified as negative. A test with perfect discrimination (no overlap in the diseased and non-diseased distributions: no false positives or false negatives) has a ROC curve that passes through the upper left corner of the ROC graph, providing 100% sensitivity and 100% specificity, but this is very rare. Therefore, the closer the ROC curve I to the upper left corner of the graph, the higher the overall accuracy of the test (21). In the dog study (14), the owner's impression of change (improved or unchanged) and the corresponding change in HRQL score were used to calculate a series of sensitivity and 1—specificity value pairs, which then made up the points on the ROC curve. Each point on the ROC curve was translated back to a value: a change in score. A point on the curve was chosen as the MID with due regard given to the consequences of the clinical implications of that choice.

The aim of this paper was to implement similar methods to those employed for the dog to improve the interpretability of the cat tool through normalization of scores, creation of a health status threshold and calculation of the MID.

STUDY 1: NORMALIZATION AND CREATION OF A THRESHOLD

Materials and Methods

Data

HRQL data for 107 healthy cats and 95 sick cats (**Supplementary Material 3**) were collected from University of Glasgow Small Animal Hospital (UGSAH), five general veterinary practices and one feline medicine specialist veterinary clinic, as part of previous HRQL instrument development studies, for which ethical approval was granted by the University of Glasgow Veterinary School. In the sick group there were no exclusion criteria and the only inclusion criterion was that the cats were suffering from a non-acute condition deemed likely to affect their QOL by the attending veterinary surgeon. The healthy cat group comprised cats deemed to be healthy by the attending veterinary surgeon.

Normalization Process Using Healthy Cats

Using raw owner-generated data from the 107 healthy cats (**Supplementary Material 2**), normalization was a two step process as follows.

Step 1: Transformation of raw 0−6 scores for each of the three HRQL domains to a continuous scale, using a logit transformation.

To allow the use of a logit transformation, HRQL scores (d) on the scale of 0−6 were converted to lie between 0 and 1, excluding exact 0 and 1 values. This was achieved by adding an arbitrary value, 0.1, at each end of the scale and dividing by 6.2 (the maximum score after the arbitrary values has been plus 0.1) as follows: $d' = \frac{d+0.1}{6.2}$.

Thereafter these converted scores were *logit* transformed to the continuous real scale (values between very large negative and very large positive values) as follows: $d'' = logit(d') = \log\left(\frac{d'}{1-d'}\right)$. The transformation to a *logit* scale puts the

TABLE 1 | Descriptive statistics for all three HRQL domain scores both raw (A) and normalized (B), for questionnaires 1−5 completed by owners of a group of 21 overweight cats enrolled on a weight loss programme.

Domain	Questionnaire No	Mean	SD	Minimum	Q1	Median	Q3	Maximum
(A)								
Vitality—Raw score	1	2.70	0.62	1.33	2.26	2.74	3.19	3.82
	2	2.91	0.53	1.50	2.60	3.05	3.30	3.54
	3	2.98	0.54	1.87	2.62	3.02	3.35	3.95
	4	3.30	0.34	2.46	3.09	3.33	3.54	3.78
	5	3.61	0.43	2.42	3.32	3.81	3.89	3.95
Comfort–Raw score	1	5.57	0.35	4.58	5.34	5.57	5.87	6.00
	2	5.61	0.54	3.97	5.54	5.78	5.94	6.00
	3	5.71	0.38	4.82	5.59	5.84	6.00	6.00
	4	5.83	0.28	4.87	5.78	5.94	6.00	6.00
	5	5.91	0.19	5.34	5.91	6.00	6.00	6.00
EWB—Raw score	1	3.40	0.50	2.47	3.05	3.38	3.99	3.99
	2	3.44	0.46	2.30	3.18	3.39	3.87	3.99
	3	3.55	0.35	2.74	3.33	3.54	3.86	3.99
	4	3.61	0.31	3.02	3.37	3.69	3.78	3.99
	5	3.75	0.27	3.06	3.59	3.75	3.99	3.99
(B)								
Vitality—Normalized	1	46.08	8.94	24.85	40.09	46.83	52.97	62.01
	2	49.10	7.64	28.00	44.96	51.14	54.57	57.98
	3	50.15	7.58	34.29	45.18	50.79	55.25	63.99
	4	54.60	4.73	43.03	51.62	54.96	57.81	61.41
	5	59.06	6.01	42.39	54.82	61.89	63.09	63.99
Comfort–Normalized	1	45.65	8.29	31.49	39.38	43.24	51.51	59.55
	2	47.67	9.23	27.00	42.61	48.28	55.04	59.55
	3	49.86	8.85	33.62	41.48	50.47	59.55	59.55
	4	52.97	7.75	34.01	48.25	54.82	59.55	59.55
	5	55.69	6.38	39.37	53.45	59.55	59.55	59.55
EWB—Normalized	1	46.28	10.45	27.14	38.98	45.73	58.84	58.84
	2	47.11	9.57	23.40	41.62	46.06	56.26	58.84
	3	49.35	7.43	32.74	44.72	49.19	55.89	58.84
	4	50.60	6.55	38.32	45.56	52.38	54.31	58.84
	5	53.62	5.79	39.31	50.13	53.71	58.84	58.84

measurement on a continuous scale, which is standard statistical practice in the testing literature (22).

Step 2: Following transformation to a continuous distribution, T-scores were calculated for 3 HRQL domain scores based on the sample means (μ), and sample SDs (σ), of the scores as follows: $T = \frac{d'' - \mu}{\sigma}$.

The T-scores have mean 0 and standard deviation 1. Finally, T-scores were scaled by multiplying them by 10 and adding 50, thus providing easily interpretable scores, where a score of 50 represented the healthy population norm for a given HRQL domain $S = 10 \times T + 50$.

In combination, these steps produce scores that are comparable across all HRQL domains, with the distributions for each set of scores having the same mean (50) and standard deviation (10).

Creation of a Threshold to Distinguish Healthy From Sick Cats

To allow the choice of a threshold which was consistent across all HRQL domains, with each domain having the same mean and standard deviation, a density plot of the theoretical distribution of the healthy cat population was constructed and the HRQL score representing the 30th percentile identified. Thereafter boxplots of the normalized HRQL domain scores (Vitality, Comfort and Emotional Well-being) for sick and healthy cats were constructed and examined visually to establish if the chosen threshold was effective at separating sick and healthy populations.

Examples Used to Demonstrate the Practical Application of Score Normalization and Threshold Creation

Examples were drawn from 3 previous studies carried out using the HRQL scale.

Example 1 illustrates the raw and normalized scores for a group of overweight cats enrolled in a Feline Weight Management Programme. Example 2 shows three groups of osteoarthritic cats, each with different severity of disease. The third example is the HRQL profile of an elderly, un-well cat whose HRQL was recorded over time, specifically to facilitate end of life discussion between owner and veterinary clinician.

Results

Data

The mean age +/– standard deviation of the cats was 6.9 +/– 2.96 (range <1–11 years) and 11.1 +/– 4.25 (range <1–20 years) for healthy and sick cats, respectively. There was no significant difference between the groups (Two–Sample T–Test, $p = < 0.001$). There were 55 males and 52 females in the healthy group and 47 males and 48 females in the sick group. The majority of cats were domestic shorthair.

Normalization Process Using Healthy Cats

The left panel of **Figure 1** shows the raw score distribution for the sick (red) and healthy (green) cats. There is considerable overlap in the Vitality and EWB domains, but less so in the Comfort domain. In the Comfort domain the boxplot for the healthy cats is non-symmetrical, with no tail on the right-hand side. This is clear evidence of a ceiling effect in the healthy cat group, with a high percentage of owners (42 of 107) recording a maximum score of 6. The right panel shows the scores once normalized. Now, because all scores are presented relative to the average healthy dog (score of 50), the domains are directly comparable to each other.

Establishing Thresholds Between Healthy and Sick Cats

Figure 2 shows the density plot of the theoretical distribution of the healthy cat population. Marked on the plot is (A), a vertical line denoting the threshold for dividing healthy and sick cats, the 30th percentile (a HRQL score of 44.8). Above this threshold (to the right of the plot) are 70% of the healthy cat population predicted to be healthy, the area marked by B. Below the threshold (to the left of the plot) are 30% of the healthy population predicted to be sick, the area marked by C. The right panel of **Figure 1** shows the normalized scores for healthy and sick cats with the shaded gray area representing the area below the 44.8 threshold score, and although the normalized scores for the sick and healthy cats overlap, this was considered acceptable at the 44.8 cut-off point.

Practical Applications

Figure 3 and **Table 1** show how the appearance of the raw and normalized scores differed in a group of 21 overweight cats enrolled in a Feline Weight Management Programme, where owners completed one assessment before treatment to provide baseline data, then four assessments at approximately monthly intervals following treatment. Formal analysis using a one-way ANOVA showed, for raw scores, $p = <0.001$ for Vitality, $p = 0.021$ for Comfort and $p = 0.043$ for EWB and for normalized scores $p = <0.001$ for Vitality, $p = 0.001$ for Comfort and $p = 0.045$ for EWB, confirming a trend for improvement over time for all domains. In the raw scores the improvement was clearest in the Vitality domain, but with the normalized scores the improvement was equally clear in all three domains. Again there is evidence of a ceiling effect in the Comfort domain where the maximum score in questionnaires 1–5 was 6 (**Table 1**). However, despite this ceiling effect, the normalized scores for each domain are much more readily interpretable, both in relation to each other and to the chosen common threshold value of 44.8.

Figure 4 demonstrates the value of having the reference scores (50 and 44.8) when interpreting the impact of different levels of disease severity, in this case in a cohort of cats with osteoarthritis (OA), classified as mild, moderate or severe by the attending clinician. The mean scores for all groups of cats were below that of the average healthy cat for all domains indicating that OA impacted the quality of life of cats even when only mildly affected by the disease.

Table 2 shows the descriptive statistics for all three HRQL domains for mild, moderate and severely affected OA cats.

Formal analysis using a one-way ANOVA showed $p = 0.631$ for Vitality, $p = 0.010$ for Comfort and $p = 0.325$ for EWB. Although only Comfort was statistically significant, the mean HRQL scores show a trend which support that the HRQL scores decline with OA severity (mild > moderate > severe).

Figure 5 illustrates the scores profile for an 18-year-old Bengal female neutered cat with OA, hyperthyroid and controlled hypertension, recorded over a 2-month period. At enrolment all three domain scores were below the healthy cat average. There was a steady decline in Emotional Well-being scores over the period, but scores for Comfort and Vitality domains were stable until week 5, after which they declined, Vitality showing more deterioration than Comfort.

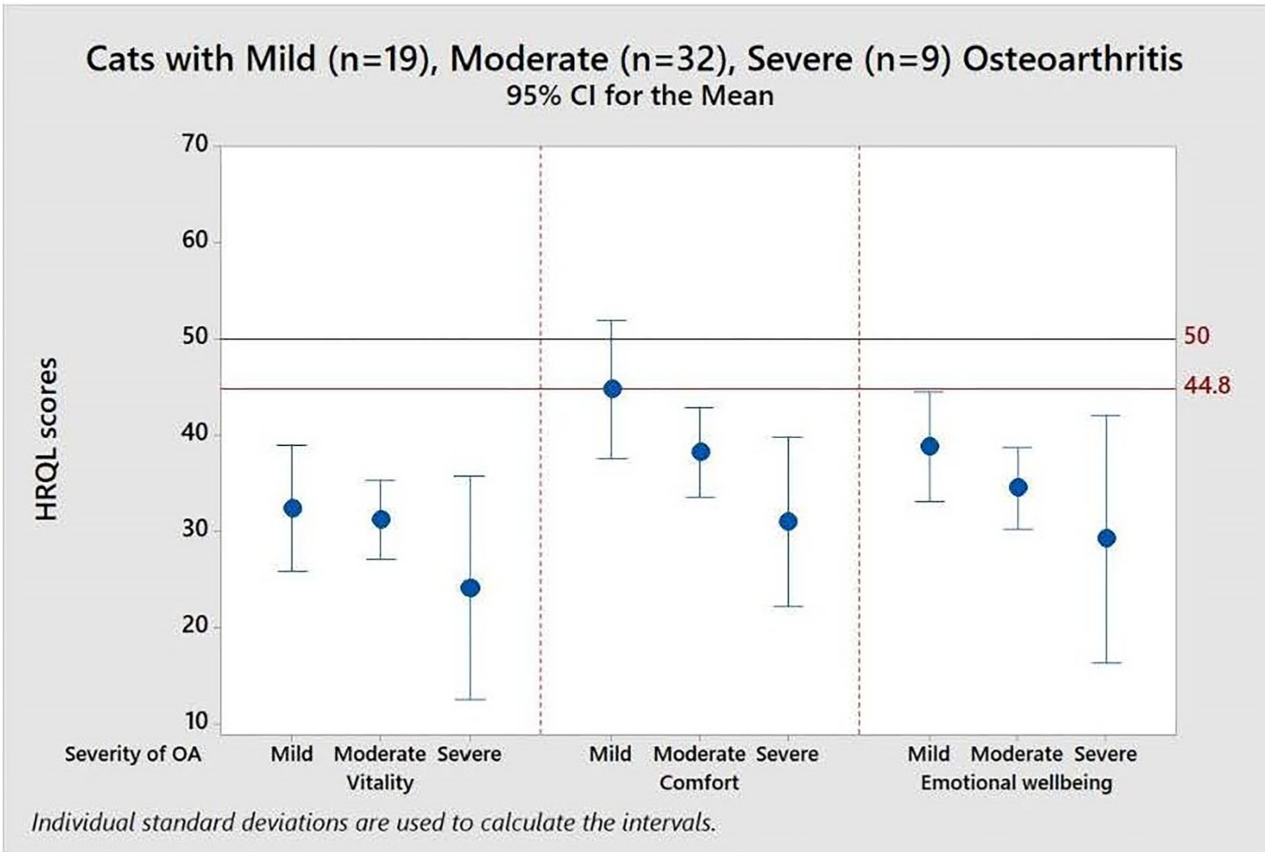


FIGURE 4 | Means and 95% confidence intervals for HRQL scores in Vitality, Comfort, Emotional Well-being for a group of 60 cats with varying severity (mild, moderate, severe) osteoarthritis. A score of 50 = the average healthy cat score and 70% of healthy cats will score above the 44.8 threshold.

TABLE 2 | Descriptive statistics for all three HRQL domains for mild, moderate and severely affected OA cats.

Domain	Severity	Mean	SD	Minimum	Q1	Median	Q3	Maximum
Vitality	Mild (*n* = 19)	32.35	13.68	11.86	20.89	30.14	43.01	56.10
	Moderate (*n* = 32)	31.19	11.34	9.20	24.01	31.60	39.91	52.03
	Severe (*n* = 9)	24.06	15.06	0.05	15.31	20.31	37.00	49.38
Comfort	Mild (*n* = 19)	44.75	14.90	22.02	31.90	44.80	57.01	70
	Moderate (*n* = 32)	38.12	12.92	23.26	30.38	34.10	41.61	70
	Severe (*n* = 9)	30.89	11.43	15.71	22.76	26.10	43.61	47.56
EWB	Mild (*n* = 19)	38.72	11.83	14.00	29.50	38.80	48.71	58.80
	Moderate (*n* = 32)	34.43	11.81	0.31	30.05	35.90	43.56	52.32
	Severe (*n* = 9)	29.16	16.78	0.00	17.94	29.14	40.39	58.80

STUDY 2: CALCULATING THE MID FOR AN IMPROVEMENT IN THE NORMALIZED HRQL SCORES

Materials and Methods

Data

Data were obtained from previous studies as before for a mixed (healthy and unhealthy) group of 95 cats (**Supplementary Material 4**) with HRQL scores collected on two or more occasions from Glasgow University Small Animal Hospital, Liverpool University Small animal Practice, five veterinary practices in the UK, one in Canada and Australia and three in the US. In addition to the 20 questions comprising the feline VetMetrica assessment, owners were asked to record whether they believed their cat's health had improved, stayed unchanged or worsened since the previous assessment. Of the 95 cats, data from two cats was incomplete, owners of 29 cats considered health status had improved, 58 had not seen a change and six considered their cat's health had deteriorated. These six cats were removed based on their small number. The unchanged group of 58 cats contained cats that were both healthy and unhealthy.

Calculation of Possible MIDs

Only the 1st and 2nd second assessments from each cat were used. For each cat, the difference between the normalized scores for assessments 1 and 2 for each domain was calculated, and generally these followed a normal distribution. The mid-point between the difference in normalized scores for each domain was selected and using the corresponding owner impression of change (unchanged or improved health) the sensitivity and specificity for each possible MID (mid-point) was calculated. For each HRQL domain, the sensitivities were then plotted against the corresponding 1—specificities to create the ROC curves for each HRQL domain (14).

Selection of MIDs

The methodology is reported by Davies et al. (14) and can be found reproduced with permission in the **Supplementary Material 1**. Briefly, using ROC curves and all possible calculated MIDs, several different methods to calculate the most appropriate MID for each HRQL domain were considered. A ROCconsistent method as described by Davies et al. (14), where all domains have the same MID followed by a similar method which allowed for different MIDs in different domains (ROCdomain) was used. Finally a set of MIDs (VetMetrica Cat) were chosen based on their position on the ROC curve.

Results

Table 3 shows the sensitivity, specificity and classification accuracy for ROCconsistent, ROCdomain and VetMetrica Cat methods. **Figure 6** shows the ROC curves constructed for

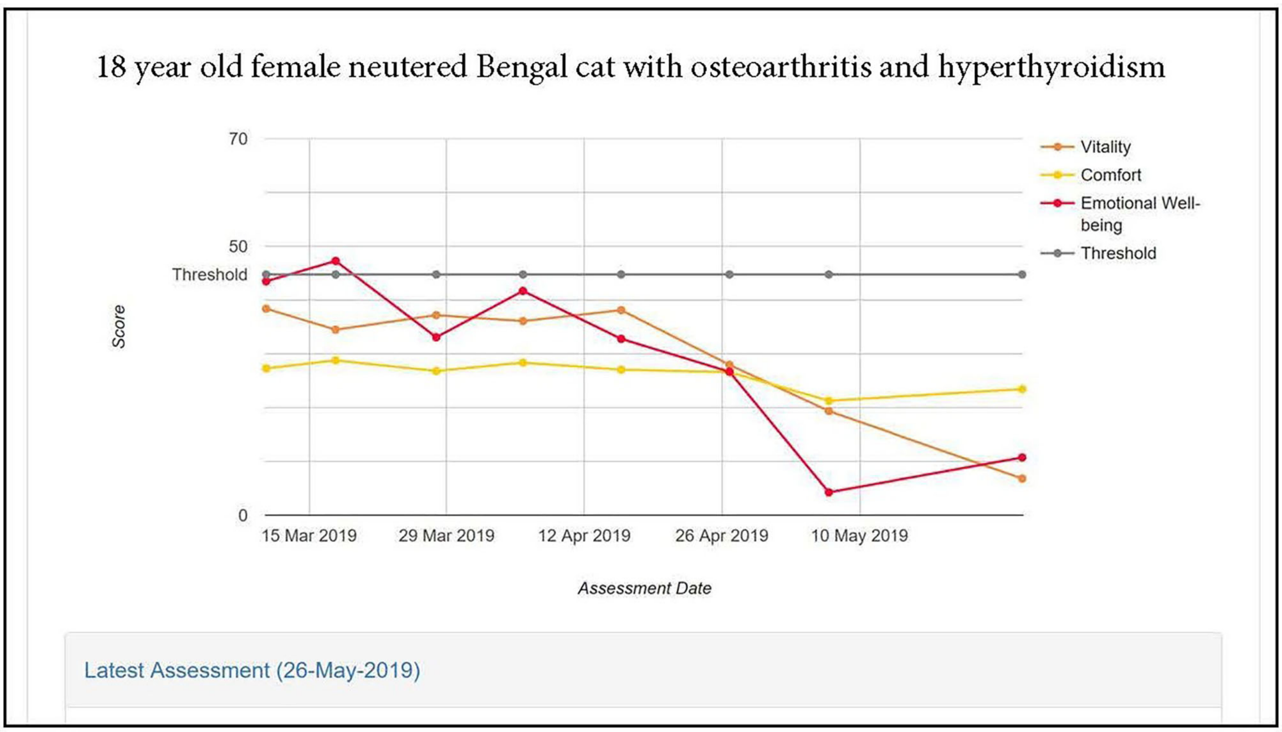


FIGURE 5 | HRQL scores in Vitality, Comfort and Emotional Well-being for an 18 year old female neutered Bengal cat with osteoarthritis and hyperthyroidism. A score of 50 = the average healthy cat score and 70% of healthy cats will score above the 44.8 threshold.

each HRQL domain as well as the corresponding sensitivities and 1—specificities. The ROC curves show the sensitivity and specificity trade off that must be considered when choosing the final MIDs. After considering the clinical implications from the different options displayed in **Table 3** and **Figure 6**, the chosen MID values for VetMetrica Cat were 5, 7.5, and 5, respectively for the Vitality, Comfort and EWB HRQL domains. **Figure 7** shows boxplots of the change in normalized HRQL domains scores. For each domain, separate boxplots are given for cats whose owners reported an improvement in health (green) and no change in health (red). On each boxplot the area below the MIDs are shaded, with the MIDs being 5, 7.5, and 5, respectively for the Vitality, Comfort and EWB HRQL domains. In each case the boxplots show an acceptable demarcation between the cats who have improved in health and those that have remained unchanged.

TABLE 3 | The sensitivities, specificities and classification accuracies of MIDconsistent, MIDdomain and VetMetrica Cat methods of MID calculation for HRQL domains Vitality, Comfort and Emotional well-being (EWB).

Domain	Method	MID	Sensitivity	Specificity	Accuracy
Vitality	MIDconsistent	10.8	0.40	0.97	0.77
Vitality	MIDdomain	10.8	0.40	0.97	0.77
Vitality	VetMetrica Cat	5	0.60	0.72	0.68
Comfort	MIDconsistent	10.8	0.50	0.81	0.70
Comfort	MIDdomain	15.5	0.47	0.93	0.77
Comfort	VetMetrica Cat	7.5	0.70	0.74	0.73
EWB	MIDconsistent	10.8	0.23	1	0.74
EWB	MIDdomain	5.4	0.60	0.90	0.80
EWB	VetMetrica Cat	5	0.60	0.86	0.77

DISCUSSION

Previously we have described three strategies to enhance the interpretability of a web-based, generic, profile HRQL measure for the dog (VetMetrica) (14) and this paper describes how these, namely score normalization, health status threshold and MID, were applied to the feline VetMetrica instrument. Norm-based scoring algorithms transformed the raw scores such that, on a 0–100 continuous scale, 50 represented the average healthy cat with a standard deviation of 10, thus scores above 50 are better than average and those below are worse compared to the healthy population. In the dog HRQL scores were normalized to the average healthy dog, according to two age groups, 0–≤7 yrs and ≥8 yrs. However, for the cat there were insufficient data to form similar age groups. Whereas, the authors were concerned that it was unrealistic to have a 1 year old dog in the same group as one that was 7 years old (14), subsequent work investigating the impact of age, breed and sex on QOL has shown that the decline in QOL with age in healthy dogs is very slow (in press). For example the decline in the score for Energetic/Enthusiastic over a 12 month period was 0.05. Clearly it would be inappropriate to extrapolate these findings to the cat, but in general it may not be as important to divide subjects into different age groups as was once thought.

The normalization process does not take account of the ceiling effects which occur when scores reach the maximum as a result of high numbers of healthy subjects scoring very highly. In the dog ceiling effects were seen in "Active/Comfortable" and

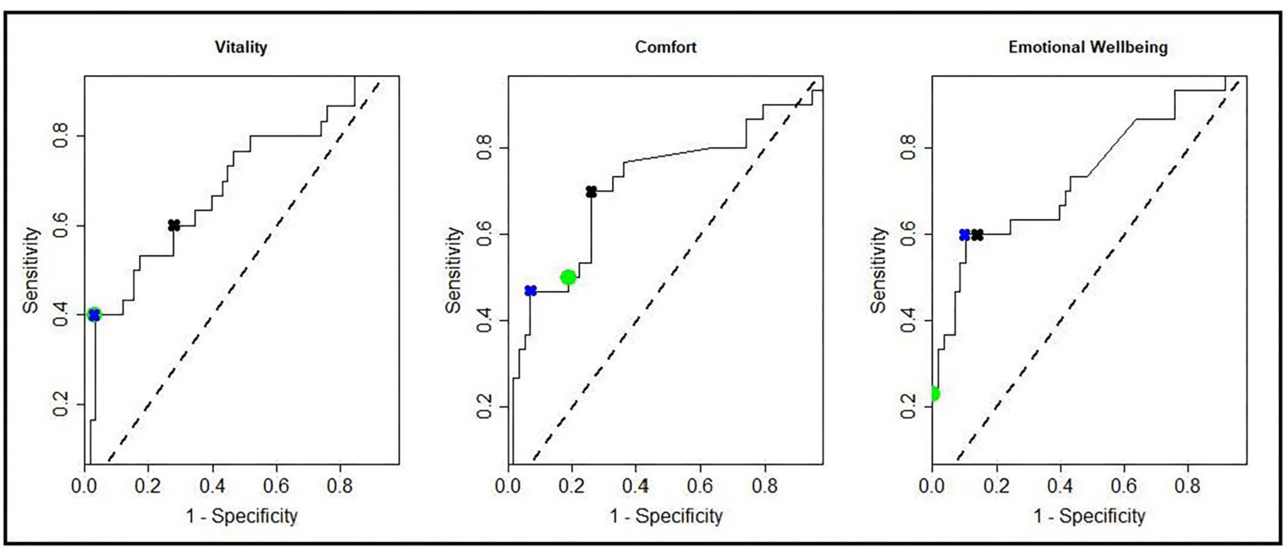


FIGURE 6 | ROC curves showing all possible MIDS for each of the HRQL domains; Vitality, Comfort, Emotional Well-being. Marked on the ROC curves are the sensitivities and 1—specificities for the MIDconsistent method (green circles), MIDdomain method (blue crosses), and the MIDs used in the cat section of the VetMetrica application (black crosses).

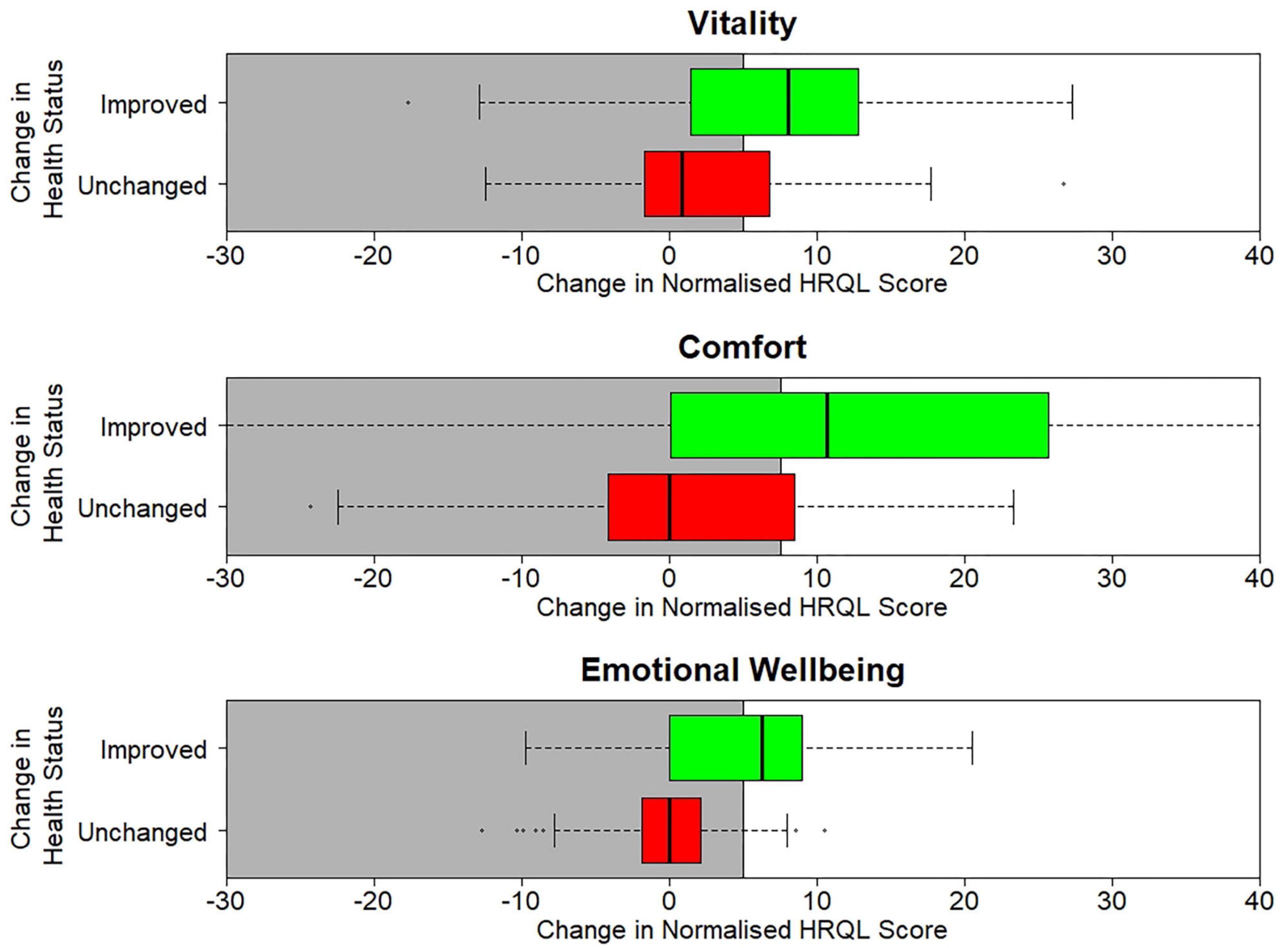


FIGURE 7 | Boxplots of the change in normalized scores in each of the three normalized HRQL domains, Vitality, Comfort and Emotional Well-being. For each domain there are two boxplots, one for cats whose health is unchanged (red) and one where their health has improved (green).

"Happy/Content" domains which reflect physical and emotional well-being, respectively, but in the cat only occurred in the "Comfort" domain which reflects physical well-being. Many people accept that cats may appear less expressive than dogs when it comes to their emotions, and therefore the behaviors making up the EWB domain may be less overt than their equivalent in the dog. Accordingly, this may have been a contributing factor to the lack of ceiling effect in the EWB domain.

Several methods, both parametric and non-parametric, have been proposed to deal with ceiling effects (23, 24), but due to the large number of cats that achieved a maximum score in the "Comfort" domain it was considered that these would not be effective at correcting the skewness of these data. Accordingly, it was decided to follow the standard statistical practice of transforming data to a continuous scale (22), and then calculating norm-based scores that are comparable across all the domains (25). It is important to note that the difference between normalized scores generates approximately normally distributed data, so ceiling effects do not affect the calculation of a MID.

A comparison of the raw score profile for a group of overweight cats recruited to a weight loss programme with the normalized profile (**Figure 3**) demonstrates the superior interpretability of normalized scores. Whereas, there is no reference point for the raw scores, the score of 50 for the average healthy cat and the threshold of 44.8 above which 70% healthy cats will score provide a useful reference point which also allows a direct comparison between domains, which are now presented on the same metric. At baseline the HRQL of the overweight cats is below average in all three domains, although scores are above the 70% threshold. By questionnaire 3 (2 months into the weight loss programme) they have improved such that the mean of the group is equivalent to the average healthy cat. That improvement is then sustained until the end of the trial. Furthermore, the two case studies presented (**Figures 4**, **5**) demonstrate how the normalized scores and threshold provide the veterinary surgeon with an immediate visual interpretation of individual or group scores relative to health status over time. Notably, many owners of cats with OA don't recognize the signs of mild disease, and yet, at that level, the condition has a marked impact on all domains of QOL, especially Vitality and EWB (**Figure 4**). In the individual cat (**Figure 5**), the physical and emotional impacts of disease follow a different trajectory, with EWB (emotional impact) declining steadily from the start of screening compared with Vitality and Comfort (physical impact) which remain stable before declining sharply at different time points. This ability to distinguish emotional from physical impact is one of the advantages of a profile measure compared with a single item score which only tells us whether a patient is better or worse, but no more. On a day to day basis there will be natural variation in domain scores and so it is important to distinguish clinically significant change from "noise." It is important to be able to advise owners of healthy cats that some change is within normal parameters, and unlikely to indicate any health change. Conversely, the clinician needs to know when an improvement in scores represents a positive effect of treatment in the sick cat. This is the function of the MID. It is equally important to determine if a deterioration in scores is meaningful, but unfortunately there were insufficient data available to investigate this. Several factors may have contributed to this. Owners may have habituated to their cat's condition and believed them to be unchanged. Alternatively they may not have remembered their cat's previous health status accurately (recall bias). In any study, this bias can be more significant when the participant has a poor memory in general and when the interval between events is long. Other factors that can influence it include age, education, socioeconomic status and the importance of the outcome to the respondent. Regarding the latter, owners may be reluctant to admit their cat has deteriorated (social desirability bias) and so it may be that some "unchanged" cats were in fact worse. The authors accept that this is a significant limitation to the use of the scale which will be addressed as more data become available. The sick cats used in this study were suffering from a variety of chronic diseases expected to impact their QOL which was appropriate for a generic scale. However, it is important to appreciate that, like validity, the MID is not an inherent property of the scale, but a feature of the scale as it is used in a particular clinical context. Accordingly, the MIDs of 5, 7.5, and 5 calculated for Vitality, Comfort and EWB, respectively, in this study for a general population may not apply when the scale is used in disease specific populations.

Small sample sizes limited the scope of these studies. While the available data were considered adequate for the normalization to the healthy cat population, it may be that experience with the tool will demonstrate that age should also be incorporated in the normalization process, as it was in the dog. Furthermore, the lack of a MID for deterioration as well as improvement is a limitation to its use. However, it is not uncommon for existing tools to undergo a continual process of refinement to accommodate new populations and contexts in which they are to be used (26).

In conclusion, if a measurement instrument is not easily interpreted, it is of limited use in clinical practice and research. This work substantially improves the interpretability of the VetMetrica generic HRQL instrument for the cat and will contribute to the body of knowledge regarding the impact of chronic disease on the emotional and physical health of this enigmatic species.

DATA AVAILABILITY STATEMENT

The original contributions generated for this study are included in the article/**Supplementary Materials**, further enquiries can be directed to the corresponding author.

ETHICS STATEMENT

The original animal study was approved by the University of Glasgow Ethics and Welfare Committee and this retrospective data analysis by the RCVS Ethics Review Panel. Written informed consent was obtained from the owners for the participation of their animals in this study.

AUTHOR CONTRIBUTIONS

VD, JR, and MS: conceptualization, writing, review, and editing. JR: data curation and original manuscript. VD and MS: statistical analysis. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

The authors wish to thank all the cat owners, as well as the veterinary surgeons and nurses in practice and at the University of Glasgow Small Animal Hospital, who willingly participated in the original studies from which we obtained data and Dr. Barr Hadar, Ontario Veterinary College, University of Guelph for use of his weight loss trial data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2020.601304/full#supplementary-material

Supplementary Material 1 | Calculating the MID for improvements in the normalised scores.

Supplementary Material 2 | Profile of scores for 107 healthy cats.

Supplementary Material 3 | Profile of scores for 202 sick and healthy cats.

Supplementary Material 4 | Profile of scores for the 1st and 2nd assessments from 95 sick and healthy cats.

REFERENCES

1. Gyatt GH, Naylor CD, Juniper E, Heyland DK, Jaeschke R, Cook DJ. User's guides to the medical literature. XII. How to use articles about health-related quality of life. *J Am Med Assoc*. (2007) 277:1232–7. doi: 10.1001/jama.1997.03540390062037
2. Spofford N, Lefebvre SL, McCune S, Niel L. Should the veterinary profession invest in developing methods to assess quality of life in healthy dogs and cats?. *J Am Vet Med Assoc*. (2013) 243:952–6. doi: 10.2460/javma.243.7.952
3. Budke CM, Levine JM, Kerwin SC, Levine GJ, Hettlich BF, Slater MR. Evaluation of a questionnaire for obtaining owner-perceived, weighted

quality- of-life assessments for dogs with spinal cord injuries. *J Am Vet Med Assoc.* (2008) 223:925–30. doi: 10.2460/javma.233.6.925
4. Favrot C, Linek M, Mueller R, Zini E. International task force on canine atopic dermatitis. Development of a questionnaire to assess the impact of atopic dermatitis on health-related quality of life of affected dogs and their owners. *Vet Dermatol.* (2010) 21:64–70. doi: 10.1111/j.1365-3164.2009.00781.x
5. Freeman LM, Rush JE, Farabaugh AE, Must A. Development and evaluation of a questionnaire for assessing health-related quality of life in dogs with cardiac disease. *J Am Vet Med Assoc.* (2005) 226:1864–8. doi: 10.2460/javma.2005.226.1864
6. Lynch S, Savary-Bataille K, Leeuw B, Argyle DJ. Development of a questionnaire assessing health-related quality-of-life in dogs and cats with cancer. *Vet Comp Oncol.* (2011) 9:172–82. doi: 10.1111/j.1476-5829.2010.00244.x
7. Niessen SJM, Powney S, Guitian J, Niessen APM, Pion PD, Shaw JAM, et al. Evaluation of quality-of-life tool for dogs with diabetes mellitus. *J Vet Int Med.* (2012) 26:953–61. doi: 10.1111/j.1939-1676.2012.00947.x
8. Noli C, Minafò G, Galzerano M. Quality of life of dogs with skin diseases and their owners. Part 1: development and validation of a questionnaire. *Vet Dermatol.* (2011) 22:335–43. doi: 10.1111/j.1365-3164.2010.00954.x
9. Reid J, Wiseman-Orr ML, Scott EM, Nolan AM. Development, validation and reliability of a web-based questionnaire to measure health-related quality of life in dogs. *J Small Anim Pract*. (2013) 54:227–33. doi: 10.1111/jsap.12059
10. Reid J, Wiseman-Orr L, Scott M. Shortening of an existing generic online health-related quality of life instrument for dogs. *J Small Anim Pract*. (2018) 59:334–42. doi: 10.1111/jsap.12772
11. Noble CE, Wiseman-Orr LM, Scott ME, Nolan AM, Reid J. Development, initial validation and reliability testing of a web-based, generic feline health-related quality-of-life instrument. *J Feline Med Surg.* (2019) 21:84–94. doi: 10.1177/1098612X18758176
12. Schünemann HJ, Akl EA, Guyatt GH. Interpreting the results of patient reported outcome measures in clinical trials: the clinician's perspective. *Health Qual Life Outcomes*. (2006) 4:62. doi: 10.1186/1477-7525-4-62
13. Lam CL, Lauder IJ, Lam TP, Gandek B. *Population Based Norming of the Chinese (HK) Version of the SF-36 Health Survey*. Hong Kong Practitioner (1999).
14. Davies V, Reid J, Wiseman-Orr ML, Scott EM. Optimising outputs from a validated online instrument to measure health-related quality of life (HRQL) in dogs. *PLoS ONE*. (2019) 14:e0221869. doi: 10.1371/journal.pone.0221869
15. Schünemann HJ, Puhan M, Goldstein R, Jaeschke R, Guyatt GH. Measurement properties and interpretability of the chronic respiratory disease questionnaire (CRQ). *J Chron Obstructive Pulm Dis*. (2005) 2:81–9. doi: 10.1081/COPD-200050651
16. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care.* (1989) 27:S178–89. doi: 10.1097/00005650-198903001-00015
17. Liang MH, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. *Med Care*. (1990) 28:632–42. doi: 10.1097/00005650-199007000-00008
18. Guyatt G, Walter S, Norman G. Measuring change over time:assessing the usefulness of evaluative instruments. *J Chron Dis.* (1987) 40:171–8. doi: 10.1016/0021-9681(87)90069-5
19. Lydick E, Epstein RS. Interpretation of quality of life changes. *Qual Life Res.* (1993) 2:221–6. doi: 10.1007/BF00435226
20. Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. *J Chron Dis.* (1986) 39:897–906. doi: 10.1016/0021-9681(86)90038-X
21. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a Fundamental evaluation tool in clinical medicine. *Clin Chem.* (1993) 39:561–77. doi: 10.1093/clinchem/39.4.561
22. Lesaffre E, Rizopoulos D, Tsonaka R. The logistic transform for Bounded outcome scores. *Biostatistics*. (2006) 8:72–85. doi: 10.1093/biostatistics/kxj034
23. Huang IC, Frangakis C, Atkinson MJ, Willke RJ, Leite WL, Vogel, et al. Addressing ceiling effects in health status measures: a comparison of techniques applied to measures for people with HIV disease. *Health Serv Res.* (2008) 43:327–39. doi: 10.1111/j.1475-6773.2007.00745.x
24. French B, Sycamore NJ, McGlashan HL, Blanchard CC, Holmes NP. Ceiling effects in the Movement Assessment Battery for Children-2 (MABC-2) suggest that non-parametric scoring methods are required. *PLoS ONE*. (2018) 13:e0198426. doi: 10.1371/journal.pone.0198426
25. Ware J, Kosinski M, Bjorner J, Turner-Bowker D, Gandek B, Maruish M. *User's Manual for the SF-36v2® Health Survey*. Lincoln (RI): QualityMetric Incorporated (2007).
26. Streiner DL, Norman GR, Cairney J. Health measurement scales: a practical guide to their development and use. USA: Oxford University Press (2015). doi: 10.1093/med/9780199685219.003.0001

Conflict of Interest: JR was employed by the company NewMetrica Ltd.

The remaining 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.

frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 15 January 2021
doi: 10.3389/fvets.2020.622837

Assessment of Rumen Mucosa, Lung, and Liver Lesions at Slaughter as Benchmarking Tool for the Improvement of Finishing Beef Cattle Health and Welfare

Luisa Magrin, Marta Brscic, Isabella Lora, Paola Prevedello, Barbara Contiero, Giulio Cozzi* and Flaviana Gottardo

Department of Animal Medicine, Production and Health, University of Padova, Padova, Italy

OPEN ACCESS

Edited by:
Edward Narayan, The University of Queensland, Australia

Reviewed by:
Rachel Allavena, The University of Queensland, Australia
Julia Adriana Calderon Diaz, Moorepark Animal and Grassland Research Centre, Teagasc, Ireland

***Correspondence:**
Giulio Cozzi
giulio.cozzi@unipd.it

Specialty section:
This article was submitted to Animal Behavior and Welfare, a section of the journal Frontiers in Veterinary Science

Received: *29 October 2020*
Accepted: *07 December 2020*
Published: *15 January 2021*

Citation:
Magrin L, Brscic M, Lora I, Prevedello P, Contiero B, Cozzi G and Gottardo F (2021) Assessment of Rumen Mucosa, Lung, and Liver Lesions at Slaughter as Benchmarking Tool for the Improvement of Finishing Beef Cattle Health and Welfare. Front. Vet. Sci. 7:622837. doi: 10.3389/fvets.2020.622837

Abattoir *post-mortem* inspections offer a useful tool for animal disease surveillance. The present cross-sectional study aimed at assessing the prevalence of rumen mucosa, lung, and liver lesions in 153 randomly selected batches of finishing beef cattle through a *post-mortem* inspection at the abattoir. At least 15 animals per batch were inspected at slaughter by two veterinarians for a total of 2,161 animals (1,376 bulls; 785 heifers) coming from 80 Italian commercial farms. Rumens were inspected by recording as binary variables (presence/absence) signs of hyperkeratosis, ruminitis, ulcer, and star scars. Similarly, lungs were inspected for signs of pneumonia and livers for signs of lipidosis, abscesses, and/or adherence. Hyperkeratosis of the mucosa and signs of ruminitis were detected in 58 and 30% of the inspected rumens, respectively. Ruminal star scars were more prevalent in bulls than in heifers (18 vs. 11%; $P < 0.05$). Signs of severe pneumonia were observed in 10% of the lungs; abscess and/or adherence in 4% of the livers. Hyperkeratosis of rumen mucosa was correlated to signs of ruminitis, and signs of ruminitis were correlated to star scars. No correlations were found between hepatic lesions and any other rumen or lung disorders. The wide variability observed among batches for the prevalence of specific lesions suggested the development of a benchmarking system to provide feedback to the farm veterinarians, as these lesions can be reflective of a subclinical disease status not easy to be detected in the live animal. Quartiles of the batch prevalence of rumen, lung, and liver alterations (if $\geq$1%) were calculated as a benchmarking tool, and third quartile value was proposed as an alarm threshold for each lesion. The use of the benchmarking system could allow to allocate each inspected batch to a specific "health class." Critical batches with a prevalence above the alarm threshold for a given lesion should be reported to veterinarians of the origin farms where actions should be taken in order to identify and lower the risk factors for that specific health issue. Knowledge of *post-mortem* inspection data along with the implementation of the proposed benchmarking system should help farm veterinarians to improve herd management from a health and welfare perspective.

Keywords: beef cattle, intensive production system, rumen hyperkeratosis, star scar, pneumonia, liver disease, *post-mortem* inspection, benchmarking system

INTRODUCTION

Post-mortem assessments at slaughter in cattle (1, 2), pigs (3), and poultry (4) have been recently considered a useful tool for animal disease surveillance. Inspections at slaughter offer the advantages of monitoring a large number of animals from several batches on the same day and collecting data from different organs in a reasonable time. This type of evaluation appears particularly cost- and labor-effective for the assessment of several digestive disorders or multiorgan diseases that could remain partially undetected *in vivo*. As in the intensive beef production systems, the appearance of feeding and management disorders could have a poor clinical manifestation (5), and the outcomes of the inspection could indirectly reveal some critical aspects of the animals' management in their farm of origin. For instance, respiratory diseases that are the most commonly reported health and economic problem both in feedlots and in intensively finishing beef cattle units (6–8) are often undetected at subclinical level or underdiagnosed on farm (9). Recently, the evaluation of the signs of bovine respiratory disease at slaughter has been proposed as an effective tool to define their negative impact during the fattening cycle (8, 10).

With regard to digestive disorders, subacute rumen acidosis (SARA) is considered an important issue for both beef cattle health and farm economy (11, 12). The development of SARA has been associated with the provision of high amounts of dietary non-structural carbohydrates (13) and/or of an insufficient amount of structured fiber (14). Recent findings in dairy cattle suggested that SARA is associated with a compromised rumen barrier and hindgut epithelium function that might allow toxin translocation and bacteria migration into the bloodstream, promoting local and systemic inflammation (15, 16). However, SARA diagnosis on beef farms is still challenging, since it is not associated with specific clinical signs (17). Ruminocentesis and rumen fluid analysis are direct diagnostic tools for SARA (18), but they are invasive and labor- and time-consuming, thus limiting the number of cattle that can be diagnosed. As an alternative to the direct on-farm diagnosis, *post-mortem* evaluation at slaughter of gross pathological evidence of rumen mucosa damage and specific liver alterations could be a retrospective strategy for the monitoring of beef cattle health, without any invasive handling on live animals.

Based on a *post-mortem* data collection at the abattoir from a wide range of randomly selected beef cattle batches coming from several fattening units, the present study aimed at (1) assessing the prevalence of different rumen mucosa, lung, and liver lesions; (2) calculating the potential correlations among different alterations detected on the same organ and among alterations detected on different organs; (3) developing a benchmarking system based on the prevalence of the recorded damage to drive cattle health improvements by farm veterinarians.

MATERIALS AND METHODS

A cross-sectional study was set up to gather information about the prevalence of rumen mucosa, lung, and liver alterations in finishing beef cattle at the slaughterhouse through a *post-mortem* inspection. Data collection was carried out in three commercial cattle slaughterhouses located in Northern Italy from April 2016 until March 2017. *Post-mortem* evaluations were carried out during 30 observation days on batches of cattle that were regularly slaughtered according to the ordinary slaughterhouse planning. Each observation day lasted from 06:00 h until 13:00 h with a target of inspecting at least six batches per day. A batch was considered a group of finished beef cattle of the same breed [Charolaise (CH), Limousine (LIM), or crossbreds (CR)] and category (bulls or heifers) coming from the same farm and belonging to the same slaughter group (same loading, transportation, unloading, lairage time, and slaughtering process). All farms of origin were located in the Po Valley (Italy), and their distance from the slaughterhouses was $<$3 h. *A priori*, it was set to inspect the organs of at least the first 15 animals per batch for batches larger than 15 animals, and of all the animals in case of smaller batches. This approach was used to trace the organs (rumen, lungs, and liver) of the same animal by two trained veterinarians located in different areas of the slaughterhouse (dispatch and tripery). The set of measures used for the study are described below and further detailed in **Table 1**. They have been chosen for their quick applicability (about 1 min/organ) without interfering with the regular working schedule of the slaughter line.

Organ Inspection

Rumens were inspected in the tripery after their dissection from intestines, omasa, and abomasa and their opening and emptying by the slaughterhouse operators. One trained veterinarian assessed rumen mucosa directly at the slaughter line having a water pump available to rinse the organ. Macroscopic alterations such as hyperkeratosis (**Supplementary Figure 1**), signs of ruminitis (**Supplementary Figure 2**), ulcers, and star scars (**Supplementary Figure 3**) were registered as binary measures (present/absent) following the methods adopted by previous studies and detailed in **Table 1**. Whenever there was presence of rumen parasites (*Paramphistomum*), it was also recorded as binary (presence/absence).

Using a simplified version of the assessment method described by previous researchers (**Table 1**), a second trained veterinarian examined lungs and livers directly in the slaughter line in the dispatch area. The assessor was positioned between the operator who detached the pluck from the carcass and the official veterinary inspector in order to visually and tactually inspect the organs before the official veterinarian did any cut or seizure. The assessor evaluated both lungs and attributed a score to the signs of pneumonia according to the description reported in **Table 1** and **Supplementary Figure 4** recording the worst condition of both lungs.

At the liver level, signs of lipidosis and the presence of abscesses and/or adherences were recorded as binary according to the criteria adapted from previous studies and reported in **Table 1**.

TABLE 1 | The scoring system used for rumen, lung, and liver evaluation at slaughter in beef cattle.

Specific disorder	Scale	Description	References	Supplementary figure
Rumen				
Hyperkeratosis	0	Absence of the alteration	Hinders and Owen (19)	**Supplementary Figure 1**
	1	Hardened rumen papillae due to a thickening of the keratin layer, recorded after visual and tactile inspection		
Signs of ruminitis	0	Absence of the alteration	Thomson (20), Thompson et al. (21), Rezac et al. (1, 2)	**Supplementary Figure 2**
	1	Absent, rarified, or immature papillae with numerous whitish or reddish nodules of 2–3 mm		
Ulcer	0	Absence of the alteration	Thomson (20), Thompson et al. (21), Rezac et al. (1, 2)	
	1	Loss of integrity of the rumen mucosa with a severe perforation and inflammatory reaction		
Star scar	0	Absence of the alteration	Thomson (20), Thompson et al. (21), Rezac et al. (1, 2)	**Supplementary Figure 3**
	1	Star-shaped scar of lamellar keratin		
Paramphistomum	0	Absence of parasites		
	1	Presence of parasites		
Lung				
Pneumonia score	0	Healthy lung	Schneider et al. (10), Leruste et al. (22)	**Supplementary Figure 4**
	1	Minimal pneumonia = one spot (1–5 cm in diameter) of gray-red discoloration		
	2	Moderate pneumonia = one larger (>5 cm in diameter) or several small spots of gray-red discoloration with a total surface of about one lobe		
	3	Severe pneumonia = gray-red discoloration area involving more than one lobe		
Liver				
Signs of lipidosis	0	Absence of the alteration	Rezac et al. (1), Attia (12)	
	1	Rounded and enlarged ventral margins of the liver potentially indicating lipidosis		
Abscess and/or adherence	0	Absence of both alterations	Rezac et al. (1), Attia (12)	
	1	Presence of superficial abscess on liver or of fibrin adherence or presence of both, visible at inspection without resection. Adherence or other alterations on the liver surface were not resected in order to avoid contamination of the organs in case of abscesses located underneath		

Statistical Methods

The statistical analyses were carried out using SAS (9.3; Institute Inc., Cary, NC) and XLSTAT (Addinsoft, New York, NY). For binary variables, score 0 was used for the absence and score 1 for the presence of alteration. Batch was the experimental unit for all the prevalence of rumen, lung, and liver alterations. Per batch, the prevalence of rumen, lung, and liver alterations was calculated as the ratio of the number of organs with a specific score over the total number of organs inspected and expressed as a percentage. Normal distribution of the batch prevalence of all alterations was tested using the Shapiro–Wilk test.

When the batch prevalence for a given lesion resulted $\geq$1%, it was tested for the association with breed and gender. In particular, the prevalence of normally distributed data regarding lesions such as rumen hyperkeratosis, signs of ruminitis, and *Paramphistomum*, and lungs with minimal (Score 1) and moderate pneumonia (Score 2) were analyzed using a mixed model that considered breed, gender, and their interaction as fixed effects, and farm as random effect, with the Bonferroni adjustment option. Mann–Whitney test was performed to analyze the effect of gender, and Kruskal–Wallis test was used to analyze the effects of breed and gender × breed interaction for

TABLE 2 | Effects of breed, gender, and their interaction on the prevalence of specific organ lesions recorded *post-mortem* at the slaughterhouse in 153 batches of finishing beef cattle coming from 80 Italian commercial farms.

Specific disorder	Overall[1]	Breed (B)			Gender (G)		Significance		
		CH	LIM	Other	Bull	Heifer	B	G	B × G
Rumen mucosa									
Hyperkeratosis[2]	57.5	53.1 ± 3.0	58.5 ± 5.2	61.6 ± 4.3	58.9 ± 3.1	56.6 ± 4.0	ns	ns	ns
Signs of ruminitis[2]	29.9	25.9 ± 2.6	29.4 ± 4.8	35.6 ± 4.1	32.8 ± 2.7	27.9 ± 3.7	ns	ns	ns
Ulcer	0.4								
Star scar[3]	14.9	15.0 (11.1–18.8)	12.0 (7.9–16.1)	18.4 (10.9–25.9)	17.5[a] (13.7–21.3)	11.4[b] (6.9–15.9)	ns	*	ns
Lung									
Minimal pneumonia[2]	19.8	17.6 ± 1.4	21.7 ± 2.6	22.5 ± 2.3	21.9 ± 1.5	19.4 ± 2.1	ns	ns	ns
Moderate pneumonia[2]	9.3	9.3 ± 1.1	6.4 ± 1.9	11.1 ± 1.6	8.7 ± 1.1	9.2 ± 1.4	ns	ns	ns
Severe pneumonia[3]	10.3	9.7 (6.9–12.5)	12.7 (7.5–17.9)	10.2 (6.3–14.1)	10.6 (7.7–13.5)	9.9 (7.2–12.6)	ns	ns	ns
Liver									
Signs of lipidosis	0.1								
Abscess and/or adherence[3]	3.8	4.1 (2.8–5.4)	4.4 (2.3–6.4)	2.8 (1.1–4.5)	4.3 (3.1–5.6)	3.0 (1.6–4.4)	ns	ns	ns

[1]*Only disorders having an overall prevalence >1% were processed to test the effects of breed, gender, and their interaction.*
[2]*Normally distributed variables were analyzed using a mixed model to test the effects of breed, gender, and their interaction, and their data are expressed as Lsmeans ± SEM.*
[3]*Non-normally distributed variables were analyzed using non-parametric procedures to test the effects of breed, gender, and their interaction, and their data are expressed as mean and 95% confidence intervals.*
[a,b]*Values within a row with different superscripts differ (*P < 0.05).*
ns, not significant.

non-normally distributed data such as the prevalence of rumens with star scar, lungs with severe pneumonia (Score 3), and livers with abscess and/or adherence.

To find out the possible redundancies between different parameters recorded at the abattoir, correlations among the prevalence of lesions detected on the same organ or on different organs were assessed at batch level (with a batch prevalence ≥1%) using Spearman's rank correlation (PROC CORR of SAS 9.3; Institute Inc., Cary, NC).

Finally, quartiles of the batch prevalence of rumen, lung, and liver alterations (with a batch prevalence ≥1%) were calculated as a benchmarking tool, according to what was proposed by Scollo et al. (3) in pigs. The third quartile (Q3) value was proposed as an alarm threshold for each lesion, and specific quartiles were calculated within breed or gender when there was a significant effect of these factors on a given lesion.

RESULTS

Results of this study regard 2,161 animals (1,376 bulls and 785 heifers) belonging to 153 batches (97 bulls and 56 heifers) that came from 80 different fattening units (from one to seven batches/farm). The observed batches had a mean size of 29.9 ± 21.3 (SD) animals, and the average proportion of inspected animals per batch over the total number of animals within each batch was 65.5 ± 28.8%. Inspected cattle belonged to the following breeds: CH, 88 batches; LIM, 28; and CR, 37.

Almost 58 and 30% of the total inspected rumens showed hyperkeratosis or signs of ruminitis, respectively, and the prevalence of these lesions did not vary among breeds or genders (**Table 2**). Ulcers were detected only in 0.4% of the inspected rumens. The prevalence of rumens with star scars in the whole inspected sample was 15%, and it was higher for bulls than heifers (**Table 2**).

The overall prevalence of rumens with *Paramphistomum* was 37%, and it differed among breeds, being higher ($P < 0.001$) for CH (41.6 ± 2.8%) and LIM (39.4 ± 4.8%) compared to CR (21.3 ± 3.8%). It was similar for bulls and heifer (34.2 ± 2.9% vs. 34.0 ± 3.6%, respectively).

Regarding lung score distribution, the overall prevalence of lungs scored 1 (minimal pneumonia) reached almost 20% of the inspected animals. The overall prevalence of lungs scored 2 and 3 (moderate and severe pneumonia) was around 10% of the inspected animals. None of these prevalence rates differed for breed or gender effect (**Table 2**). Over the total inspected livers, those showing signs of lipidosis were very rare (<0.5%) and those showing abscess and/or adherence were almost 4%. No effects of breed or gender were found for the prevalence of liver alterations (**Table 2**).

Results of the Spearman rank correlations on lesion prevalence at the batch level are reported in **Table 3**. The prevalence of rumen mucosa with hyperkeratosis was significantly correlated ($P < 0.001$) to that of rumens with signs of ruminitis, and it was negatively correlated ($P < 0.05$) to the prevalence of lungs showing severe signs of pneumonia. A positive correlation ($P < 0.001$) was also found between the prevalence of rumens with signs of ruminitis and that of rumens with star scars (**Table 3**).

The quartiles of the prevalence of the specific lesions in the inspected batches are shown in **Table 4**. As the rumen hyperkeratosis was recorded with high frequency in most of

TABLE 3 | Correlations among different rumen mucosa, lung, and liver lesions recorded *post-mortem* at the slaughterhouse in 153 batches of finishing beef cattle coming from 80 Italian commercial farms.

Specific disorder	Rumen mucosa			Lung			Liver
	Hyperkeratosis	Signs of ruminitis	Star scar	Minimal pneumonia	Moderate pneumonia	Severe pneumonia	Abscess and/or adherence
Rumen mucosa							
Hyperkeratosis		0.284	0.054	0.130	−0.091	−0.170	−0.006
		***	ns	ns	ns	*	ns
Signs of ruminitis			0.439	0.004	0.038	−0.076	0.047
			***	ns	ns	ns	ns
Star scar				0.096	−0.151	−0.068	−0.076
				ns	ns	ns	ns
Lung							
Minimal pneumonia					−0.185	0.130	−0.099
					*	ns	ns
Moderate pneumonia						0.130	−0.100
						ns	ns
Severe pneumonia							0.062
							ns

*Different symbols indicate significant correlations (*P < 0.05 and ***P < 0.001).*
ns, not significant.

TABLE 4 | Statistical description with quartiles of the batch average prevalence of specific organ lesions recorded *post-mortem* at the slaughterhouse in 153 batches of finishing beef cattle coming from 80 Italian commercial farms.

Specific lesion (% of organs affected/batch)	Q1	Median	Q3
Rumen mucosa			
Hyperkeratosis	40.0	60.0	75.0
Signs of ruminitis	10.0	26.7	46.7
Star scar			
Bull	0.0	12.5	26.7
Heifer	0.0	3.3	15.7
Lung			
Minimal pneumonia	10.0	20.0	26.7
Moderate pneumonia	0.0	6.7	13.3
Severe pneumonia	0.0	6.7	13.3
Liver			
Abscess and/or adherence	0.0	0.0	6.7

Q1, first quartile; Q3, third quartile.

the inspected batches, its median and Q3 values were very high, followed by the signs of ruminitis with a Q3 value of 47% (**Figure 1**). Batch prevalence of star scars differed between bulls and heifers, with the former having a higher alarm threshold value than the latter (27 vs. 16%, respectively; **Figure 2**). The prevalence of minimal signs of pneumonia at the batch level was higher than those of moderate and severe signs of pneumonia, leading to different alarm threshold values (**Figure 3**). The presence of abscesses or adherence in the liver was quite limited, with an alarm threshold value around 7% (**Figure 4**).

DISCUSSION

The Italian beef cattle industry accounts for more than 1.5 million animals per year (5), nearly one third of the total beef cattle fattened in Europe. It is mainly located in the Po Valley, a fertile and homogeneous climatic area, and is based on the finishing of purebred or crossbred bulls and heifers imported from abroad at an age of 10–14 months and a live weight at the arrival of 300–400 kg (23). After their transfer to Italy, cattle are finished for about 7 months in specialized fattening units, housed in multiple pens with fully slatted or deep littered floor (24).

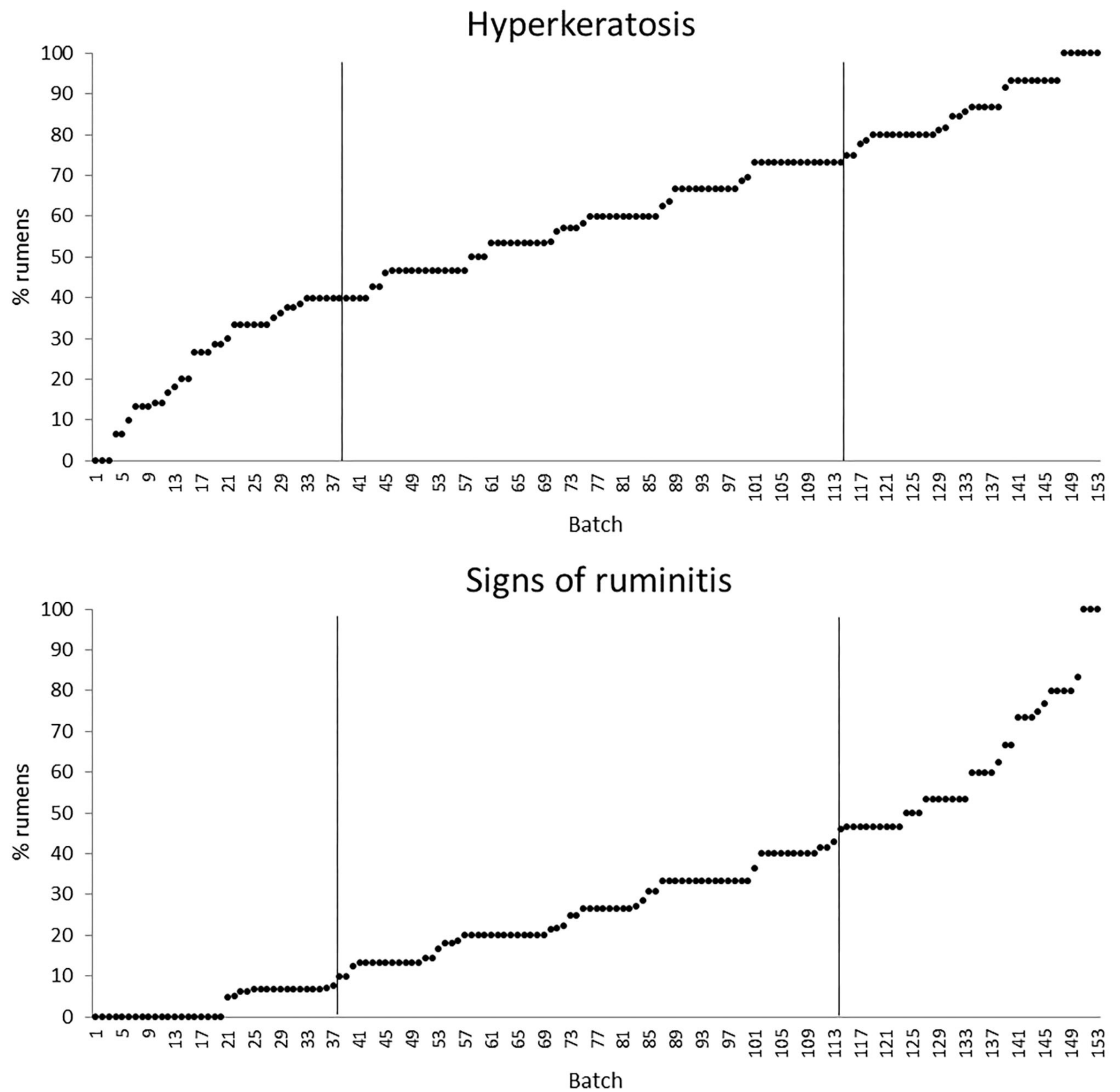


FIGURE 1 | Distribution of the batch prevalence of rumen hyperkeratosis and signs of ruminitis of 153 batches of finishing beef cattle (represented by dots) coming from 80 Italian commercial farms. Solid vertical lines indicate thresholds at first and third quartiles.

Beef farms operate according to a rather standardized feeding program made of high-starch diets based on maize silage to promote maximum daily gain (25). Both housing and feeding plans have been recognized as potential risk factors for the health and welfare of the animals (1, 11), with increased costs for medical treatments (26).

The prevalence of rumen mucosa hyperkeratosis found in this study was considerably high and so widespread among batches that might be interpreted as an adaptive response of the animals to some common challenging conditions. Indeed, this alteration of the rumen wall is mainly associated with the provision of high-starch/low-fiber diets during the finishing period (27). At rumen level, the starch load might cause a temporary imbalance between production and absorption of fatty acids with a consequent drop in ruminal pH (14, 28). When a prolonged condition of low rumen pH (5.5–5.0) persists, bacteria might invade the rumen wall and eventually lead to ruminitis and severe damage to the rumen mucosa papillae (14, 29). The quite high prevalence of

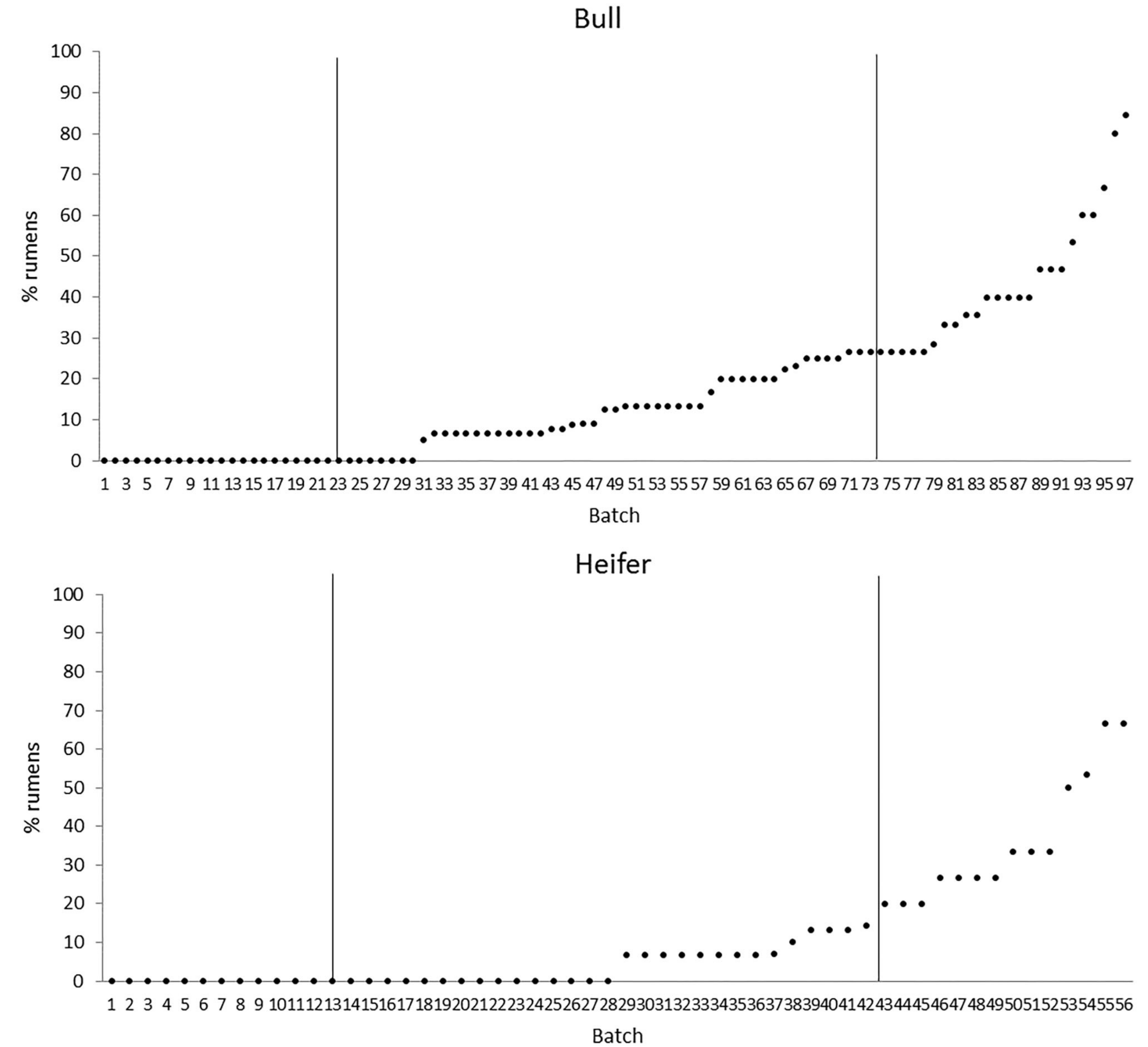


FIGURE 2 | Distribution of the batch prevalence of rumen star scars of 153 batches of finishing beef cattle (represented by dots) coming from 80 Italian commercial farms divided by gender (bull, $n = 97$ batches; heifer, $n = 56$ batches). Solid vertical lines indicate thresholds at first and third quartiles.

signs of ruminitis recorded in this study is consistent to this hypothesis. The relevant role of the dietary energy content in developing specific damage on the rumen wall has been proved by several studies on finishing feedlot cattle, where animals fed diets with more than 70% of concentrate in a finishing period of 3 months had an occurrence of rumen mucosa ulcers and/or ruminitis of 19.9–31.3% (1, 26).

Particularly alarming in the current study was the noticeable prevalence of star scars on the rumen mucosa that was similar to the values recorded in feedlot cattle (1, 26). Star scars represent the footprint of a previous pathological condition, as they are outcomes of the healing process from ruminitis or ulceration. Even their etiology arises from the ingestion of a high amount of grains or of high-starch/low-fiber rations that induce a relevant chemical insult to the ruminal mucosa with a reduction of its absorptive capacity (1, 2, 30). In the present study, the prevalence of star scars was higher for bulls than for heifers, and this result may be attributed to the different feeding plans adopted for these two cattle categories. Indeed, beef heifers are fed more roughages and fattened for a shorter period than beef bulls to prevent an excessive carcass fatness score (23).

In this study, the occasional finding of massive rumen infestation by parasites of the genus *Paramphistomum* was heavily influenced by cattle breeds likely reflecting their different

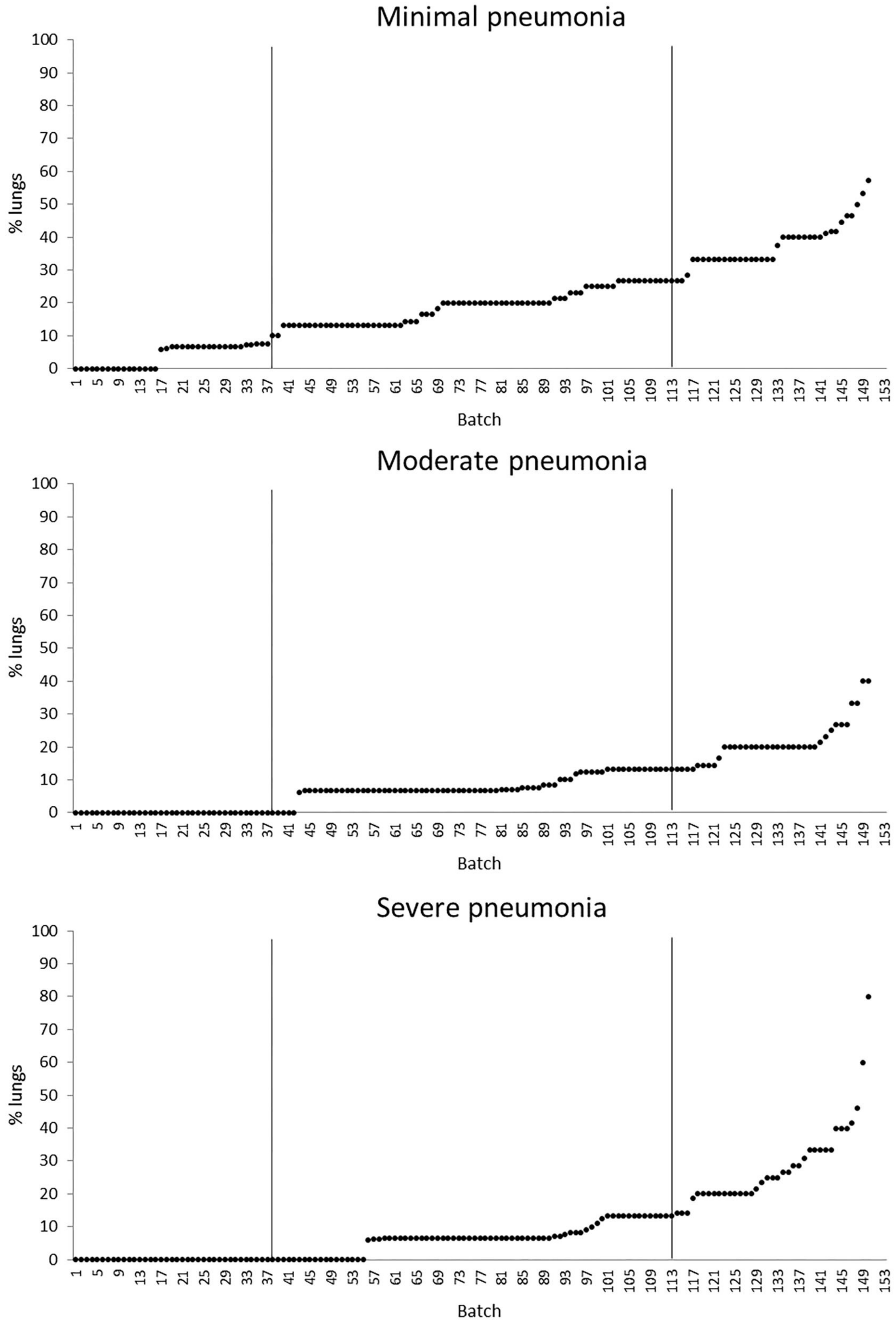


FIGURE 3 | Distribution of the batch prevalence of lung minimal, moderate, and severe pneumonia of 153 batches of finishing beef cattle (represented by dots) coming from 80 Italian commercial farms. Solid vertical lines indicate thresholds at first and third quartiles.

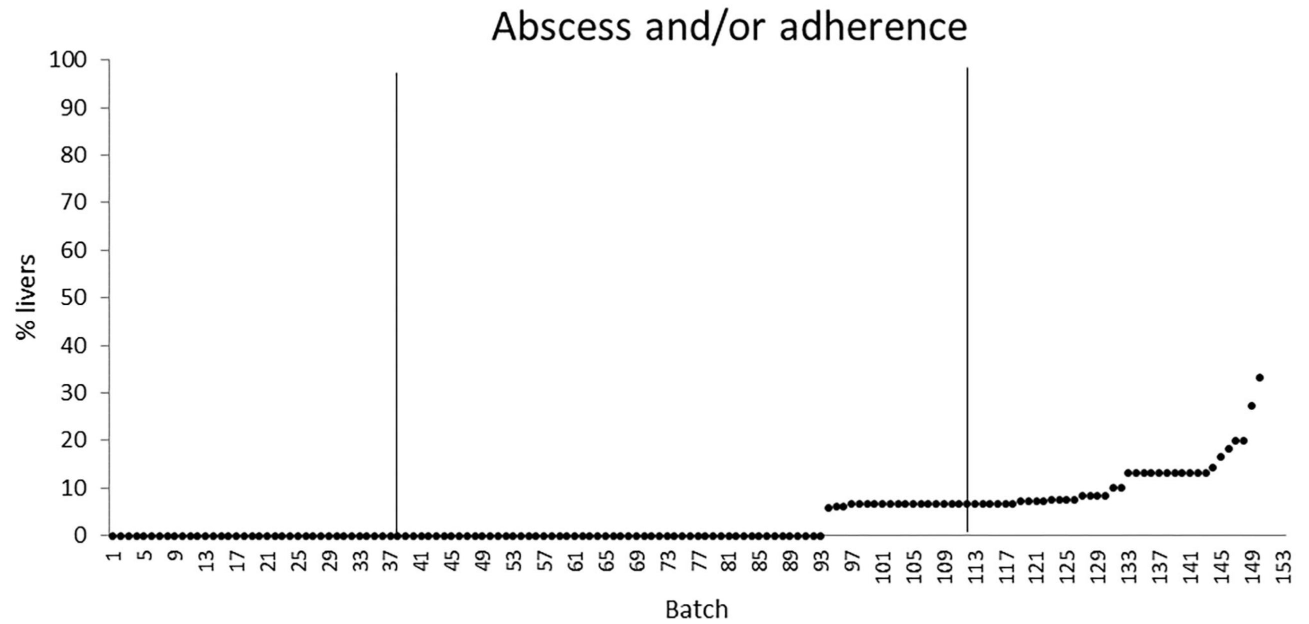


FIGURE 4 | Distribution of the batch prevalence of liver abscess and/or adherence of 153 batches of finishing beef cattle (represented by dots) coming from 80 Italian commercial farms. Solid vertical lines indicate thresholds at first and third quartiles.

origins. French CH and LIM are generally kept at pasture before the transfer to Italy and thus easily exposed to *Paramphistomum* infestation (31). Crossbred cattle are instead mainly of domestic origin, and they have a very limited or no access to pasture before finishing. *Paramphistomum* infestation usually has a subclinical course; however, the negative effect of a massive infestation on rumen health and absorptive efficiency should be carefully evaluated by further dedicated studies.

In literature, the reported prevalence of liver abscesses in feedlot cattle showed a great variability from 4.8% in McKeith et al. (32) to 13.9% in Garcia et al. (33), with the highest value of 32% reported by Nagaraja and Lechtenberg (34) for Canadian and North American cattle. The difference between these values and the prevalence recorded in the current study (about 4.0%) was presumably due to the diverse feeding regimen and environmental conditions (indoor vs. outdoor housing, climatic zone, etc.) that characterize these beef production systems. Some feedlot studies have proven the existence of an association between the occurrence of rumen pathologies and liver abscesses within the so-called ruminitis–liver abscess complex (1, 12). In our study, the low prevalence of liver abscesses and the lack of any association between them and the rumen lesions suggest that a possible SARA condition affecting the inspected batches did not induce a severe inflammatory response with the translocation of bacteria and endotoxins across the ruminal epithelium.

The overall prevalence of healthy lungs recorded in this study (on average 62.0%) was satisfactory if compared to the outcome (36%) of a similar study carried out on intensively finished beef cattle reared in the same area (8). Moreover, our percentage of healthy lungs was consistent with previous values recorded at the slaughter for feedlot cattle, which ranged from 60 to 72% (1, 10, 35). In these feedlot studies, the prevalence of clinical signs of 35 and 8.17% observed *in vivo* by Wittum et al. (35) and Schneider et al. (10), respectively, was significantly lower than that of lung lesions recorded *post-mortem*, indicating that direct *on-farm* diagnosis could neglect a large number of animals that have likely suffered from subclinical pneumonia (10). Therefore, as the evaluation of lungs at the abattoir is a reliable indicator of the real occurrence of respiratory diseases during the fattening period, data from *post-mortem* inspections should complement the *in vivo* health checks in order to detect specific risks that could impair the lungs' health on farm.

The improvement of animal health surveillance through the identification of simple reliable indicators is a priority of the EU animal health strategy (36). In this regard, *post-mortem* inspections at the abattoir are an important complement to the *intra vitam* health checks, as the recorded lesions can be reflective of subclinical disease status not easily detected in the live animal. In addition, a benchmarking system based on inspections data about the prevalence and severity of lesions at batch level could support farm veterinarians to prioritize their actions to improve the herd management from a health and welfare perspective.

Consistent with the approach proposed by Scollo et al. (3) in pigs, the calculation of the quartiles for the batch population according to the prevalence of each rumen, lung, and liver lesion allows to allocate each inspected batch in a specific "health class." The rationale behind this benchmarking system should foresee a targeted intervention by the stockman and the veterinarian in those farms for which, at slaughter, batches of cattle showed a prevalence above the defined alarm threshold (Q3) for a given lesion. The implementation of the system through a wide collection of *post-mortem* organ inspection data in several abattoirs would set reliable alarm thresholds for each lesion for a specific country or production system. At this purpose, to limit a possible bias due to the interobserver variation, it would be advisable to organize a proper training to standardize the definitions of lesions and scores across the inspectors working in different abattoirs. A regular application of the benchmarking system should promote a virtuous cycle in which beef farms where a specific intervention is needed are taking actions to improve health and feeding management and move to a better quartile. In the medium to long run, the benchmarking system should support a general and continuous improvement of beef cattle health management, resulting in better performance and a lower use of pharmaceutical treatments.

This study provided an overview of the occurrence of some specific rumen, lung, and liver lesions that affect intensively finished beef cattle. Some of the recorded lesions, like rumen hyperkeratosis and signs of ruminitis, had a very high prevalence, suggesting the need for an intervention on the feeding management during the finishing period. Knowledge of *post-mortem* organ inspection data is of value, as several lesions can be reflective of subclinical disease status not easily detected in the live animal. The variability observed among the batch prevalence for specific signs of diseases suggested the development of a benchmarking system to help farm veterinarians to drive herd health improvement. A wide implementation of this system should promote a continuous improvement of beef cattle management from a health and welfare perspective.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

Ethical review and approval was not required for the animal study because the organs' assessments were performed post-mortem only.

AUTHOR CONTRIBUTIONS

FG and GC conceived and designed study. LM, MB, IL, and PP collected, compiled, and analyzed the data. BC performed statistical analyses. LM, IL, and GC drafted and edited the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by the University of Padova within the PRAT project CPDA158107 entitled use of animal-based measures taken at the slaughterhouse as indicators of welfare at the farm: a retrospective approach.

ACKNOWLEDGMENTS

The authors wish to thank the AZOVE beef cattle producers' association and their veterinarians for their precious support and the official veterinary inspectors and all personnel of the slaughterhouses involved (Colomberotto S.p.a., Pellizzari Carni dei Fratelli Pellizzari S.n.c., and Macello Piemonte Nord S.r.l.) for hosting the research group for data collection.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2020.622837/full#supplementary-material

REFERENCES

1. Rezac DJ, Thomson DU, Bartle SJ, Osterstock JB, Prouty FL, Reinhardt CD. Prevalence, severity, and relationships of lung lesions, liver abnormalities, and rumen health scores measured at slaughter in beef cattle. *J Anim Sci.* (2014) 92:2595–602. doi: 10.2527/jas.2013-7222
2. Rezac DJ, Thomson DU, Siemens MG, Prouty FL, Reinhardt CD, Bartle SJ. A survey of gross pathologic conditions in cull cows at slaughter in the Great Lakes region of the United States. *J Dairy Sci.* (2014) 97:74227–35. doi: 10.3168/jds.2013-7636
3. Scollo A, Gottardo F, Contiero B, Mazzoni C, Leneveu P, Edwards SA. Benchmarking of pluck lesions at slaughter as a health monitoring tool for pigs slaughtered at 170 kg (heavy pigs). *Prev Vet Med.* (2017) 144:20–8. doi: 10.1016/j.prevetmed.2017.05.007
4. Louton H, Erhard M, Wöhr AC. Acquisition of animal-based welfare measures at slaughter of poultry. *Fleischwirtschaft.* (2018) 98:94–8.
5. Cozzi G. Present situation and future challenges of beef cattle production in Italy and the role of the research. *Ital J Anim Sci.* (2007) 6:389–96. doi: 10.4081/ijas.2007.1s.389
6. Thompson PN, Stone A, Schultheiss WA. Use of treatment records and lung lesion scoring to estimate the effect of respiratory disease on growth during early and late finishing periods in South African feedlot cattle. *J Anim Sci.* (2006) 84:488–98. doi: 10.2527/2006.842488x
7. Buchanan JW, Mac Neil MD, Raymond RC, McClain AR, Van Eenennaam AL. Rapid Communication: Variance component estimates for Charolais-sired fed cattle and relative economic impact of bovine respiratory disease. *J Anim Sci.* (2016) 94:5456–60. doi: 10.2527/jas.2016-1001
8. Caucci C, Di Martino G, Schiavon E, Garbo A, Soranzo E, Tripepi L, et al. Impact of bovine respiratory disease on lung lesions, slaughter performance and antimicrobial usage in French beef cattle finished in North-Eastern Italy. *Ital J Anim Sci.* (2018) 17:1065–9. doi: 10.1080/1828051X.2018.1426395
9. Edwards TA. Control methods for bovine respiratory disease for feedlot cattle. *Vet Clin North Am Food Anim Pract.* (2010) 26:273–84. doi: 10.1016/j.cvfa.2010.03.005
10. Schneider MJ, Tait Jr RG, Busby WD, Reecy JM. An evaluation of bovine respiratory disease complex in feedlot cattle: Impact on performance and carcass traits using treatment records and lung lesion scores. *J Anim Sci.* (2009) 87:1821–7. doi: 10.2527/jas.2008-1283

11. Scientific Committee on Animal Health and Animal Welfare (SCAHAW). *The Welfare of Cattle Kept for Beef Production*. Sanco.C.2/AH/R22/2000 (2001). Retrieved from: http://ec.europa.eu/food/fs/sc/scah/out54_en.pdf (accessed September 03, 2020).
12. Attia NE. Subacute ruminal acidosis in feedlot: Incidence, clinical alterations and its sequelae. *Adv Anim Vet Sci*. (2016) 4:513–7. doi: 10.14737/journal.aavs/2016/4.10.513.517
13. Hernández J, Benedito JL, Abuelo A, Castillo C. Ruminal acidosis in feedlot: from aetiology to prevention. *Sci World J*. (2014) 2014:702572. doi: 10.1155/2014/702572
14. Kleen JL, Hooijer GA, Rehage J, Noordhuizen JPTM. Subacute ruminal acidosis (SARA): a review. *J Vet Med A Physiol Pathol Clin Med*. (2003) 50:406–14. doi: 10.1046/j.1439-0442.2003.00569.x
15. Khafipour E, Plaizier JC, Aikman PC, Krause DO. Population structure of rumen *Escherichia coli* associated with subacute ruminal acidosis (SARA) in dairy cattle. *J Dairy Sci*. (2011) 94:351–60. doi: 10.3168/jds.2010-3435
16. Khiaosa-ard R, Zebeli Q. Diet-induced inflammation: From gut to metabolic organs and the consequences for the health and longevity of ruminants. *Res Vet Sci*. (2018) 120:17–27. doi: 10.1016/j.rvsc.2018.08.005
17. Nagaraja TG, Titgemeyer EC. Ruminal acidosis in beef cattle: the current microbiological and nutritional outlook. *J Dairy Sci*. (2007) 90:E17–38. doi: 10.3168/jds.2006-478
18. Gianesella M, Morgante M, Cannizzo C, Stefani A, Dalvit P, Messina V, et al. Subacute ruminal acidosis and evaluation of blood gas analysis in dairy cow. *Vet Med Int*. (2010) 2010:392371. doi: 10.4061/2010/392371
19. Hinders RG, Owen FG. Relation of ruminal parakeratosis development to volatile fatty acid absorption. *J Dairy Sci*. (1965) 48:1069–73. doi: 10.3168/jds.S0022-0302(65)88393-X
20. Thomson RG. Rumenitis in cattle. *Can Vet J*. (1967) 8:189–92.
21. Thompson PN, Schultheiss WA, Hentzen A. *The Effect of Rumen Mucosal Lesions on Growth in South African Feedlot Cattle*. Budapest: World Buiatrics Congress (2008). p. 12.
22. Leruste H, Brscic M, Heutinck LFM, Visser EK, Wolthuis-Fillerup M, Bokkers EAM, et al. The relationship between clinical signs of respiratory system disorders and lung lesions at slaughter in veal calves. *Prev Vet Med*. (2012) 105:93–100. doi: 10.1016/j.prevetmed.2012.01.015
23. Gallo L, De Marchi M, Bittante G. A survey on feedlot performance of purebred and crossbred European young bulls and heifers managed under intensive conditions in Veneto, northeast Italy. *Ital J Anim Sci*. (2014) 13:798–807. doi: 10.4081/ijas.2014.3285
24. Magrin L, Gottardo F, Brscic M, Contiero B, Cozzi G. Health, behaviour and growth performance of Charolais and Limousin bulls fattened on different types of flooring. *Animal*. (2019) 13:2603–11. doi: 10.1017/S175173111900106X
25. Cozzi G, Mazzenga A, Contiero B, Burato G. The use of corn silage in beef cattle feeding during the finishing period. *Ital J Anim Sci*. (2008) 7:39–52. doi: 10.4081/ijas.2008.39
26. Malafaia P, Granato TAL, Costa RM, de Souza VC, Costa DFA, Tokarnia CH. Major health problems and their economic impact on beef cattle under two different feedlot systems in Brazil. *Pesqui Vet Bras*. (2016) 36:837–43. doi: 10.1590/s0100-736x2016000900008
27. Meyer NF, Bryant TC. Diagnosis and management of rumen acidosis and bloat in feedlots. *Vet Clin North Am Food Anim Pract*. (2017) 33:481–98. doi: 10.1016/j.cvfa.2017.06.005
28. Plaizier JC, Khafiour E, Li S, Gozho GN, Krause DO. Subacute ruminal acidosis (SARA), endotoxins and health consequences. *Anim Feed Sci Technol*. (2012) 172:9–21. doi: 10.1016/j.anifeedsci.2011.12.004
29. Wiese BI, Campbell J, Hendrick S, Penner GB. Ruminal pH, short-chain fatty acid concentrations, and serum acute phase protein concentrations during finishing for steers with and without rumen and liver pathology. *Can J Anim Sci*. (2017) 97:581–9. doi: 10.1139/CJAS-2016-0212
30. Blowey R, Weaver AD. Alimentary disorders. In: *Color Atlas of Diseases and Disorders of Cattle*. 3rd Ed. London: Elsevier Health Sciences (2011). p. 53–82. doi: 10.1016/B978-0-7234-3602-7.00010-8
31. Sanabria REF, Romero JR. Review and update of paramphistomosis. *Helminthologia*. (2008) 45:64–8. doi: 10.2478/s11687-008-0012-5
32. McKeith RO, Gray GD, Hale DS, Kerth CR, Griffi DB, Savell JW, et al. National beef quality audit – 2011: Harvest-floor assessments of targeted characteristics that impact quality and value of cattle, carcasses, and byproducts. *J Anim Sci*. (2012) 90:5135–42. doi: 10.2527/jas.2012-5477
33. Garcia LG, Nicholson KL, Hoffman TW, Lawrence TE, Hale DS, Griffi DB, et al. National beef quality Audit – 2005: survey of targeted cattle and carcass characteristics related to quality, quantity, and value of fed steers and heifers. *J Anim Sci*. (2008) 86:3533–43. doi: 10.2527/jas.2007-0782
34. Nagaraja TG, Lechtenberg KF. Liver abscesses in feedlot cattle. *Vet Clin Food Anim*. (2007) 23:351–69. doi: 10.1016/j.cvfa.2007.05.002
35. Wittum TE, Woollen NE, Perino LJ, Littledike ET. Relationships among treatment for respiratory tract disease, pulmonary lesions evident at slaughter and rate of weight gain in feedlot cattle. *J Am Vet Med Assoc*. (1996) 209:814–8.
36. European Commission. *A New Animal Health Strategy for the European Union (2007-2013) Where Prevention is Better than Cure*. European Commission Edn. Office for Official Publications of the European Communities. Luxemburg (2007). p. 1–26.

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.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership