



A comparison of three animal welfare assessment protocols applied to Swedish dairy cow herds

B. Staaf Larsson, A. Jansson, M. Holmberg, L. Winblad von Walter, M. Stéen & K. Dahlborn

To cite this article: B. Staaf Larsson, A. Jansson, M. Holmberg, L. Winblad von Walter, M. Stéen & K. Dahlborn (21 Feb 2024): A comparison of three animal welfare assessment protocols applied to Swedish dairy cow herds, Acta Agriculturae Scandinavica, Section A — Animal Science, DOI: [10.1080/09064702.2024.2317708](https://doi.org/10.1080/09064702.2024.2317708)

To link to this article: <https://doi.org/10.1080/09064702.2024.2317708>



© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



[View supplementary material](#)



Published online: 21 Feb 2024.



[Submit your article to this journal](#)



Article views: 372



[View related articles](#)



[View Crossmark data](#)

A comparison of three animal welfare assessment protocols applied to Swedish dairy cow herds

B. Staaf Larsson ^a, A. Jansson ^a, M. Holmberg ^b, L. Winblad von Walter ^c, M. Stéen ^a and K. Dahlborn ^a

^aDepartment of Animal Biosciences, Swedish University of Agricultural Sciences, Uppsala, Sweden; ^bNational Veterinary Institute, SVA, Uppsala, Sweden; ^cVäxa Sverige, Stockholm, Sweden

ABSTRACT

This study compared the outcome of the Swedish Official Control (OC) with that of two other well-established protocols for assessing animal welfare in dairy cows, 'Ask the Cow' (AC) and Welfare Quality (WQ). Forty-one farms already scheduled for an official control were assessed by the three protocols on the same day. The hypothesis was that farms would be ranked similarly in terms of best and worst, irrespective of the protocol used. A second aim was to explore whether any of the animal-based quantified measures in AC and WQ could be candidates to improve OC. The eight farms with most remarks in OC (3–5) were ranked in the range 19–40 in AC and 5–37 in WQ. The only correlation observed ($r=0.40$, $P=0.009$) was between the rankings in AC and OC. Candidate measurements to improve OC taken from AC and WQ are quantifying individual body condition and cleanliness and recording skin lesions.

ARTICLE HISTORY

Received 11 October 2023
Accepted 7 February 2024

KEYWORDS

Animal-based measures;
animal welfare assessment;
dairy cattle; resource-based
measure

Introduction


A high level of animal protection and good animal welfare is important from an animal perspective, but also because consumers demand high-quality food produced by healthy animals kept in a sound environment (Berthe et al., 2012; KilBride et al., 2012). However, there is neither an international protocol nor a gold standard for animal welfare assessment of dairy cows, although the Welfare Quality (WQ) protocol has been used in several scientific studies carried out in different countries (e.g. Radeski et al., 2015; Van Os et al., 2018; Gieseke et al., 2022; Barry et al., 2023). Concerning housing and management of dairy cows, there are yet no specific European Union (EU) regulations in place, but general rules on keeping farm animals are included in Council Directive 98/58/EC (European Commission, 1998). In Sweden, the Animal Welfare Act (2018:1192; Swedish Government, 2018) and Ordinance (2019:66; Swedish Government, 2019) are applicable to animals kept by humans, with specific paragraphs regarding dairy cows. In addition, specific regulations concerning cattle are in place (SJVFS 2019:18, Case No L 104, Swedish Board of Agriculture, 2019).

Assessments of animal welfare are generally based on management-, resource- or animal-related parameters. The first two, also called input-based parameters, refer

to animal care and the environment influencing the animals. The outcome, i.e. how the animals are influenced by their environment, is assessed using animal-based parameters (Keeling, 2009; Radeski et al., 2015). Animal-based measures are commonly applied in animal welfare science (Keeling, 2009; Sandgren et al., 2009). In contrast, Swedish animal welfare legislation has a preventative focus (Lundmark et al., 2016), and therefore the official animal welfare control (OC) conducted by County Administration Boards (CABs) in Sweden primarily involves resource-based measures. This is in accordance with current EU legislation, which also relies on providing resources and management (Blokhuys et al., 2010).

In 2010, the Swedish Dairy Association launched an animal-based scoring system called 'Ask the Cow' (AC). This protocol is used as a benchmarking advisory tool and focuses on welfare of cows, young stock and calves. As mentioned, another protocol in use in Europe is the WQ system, which originated from an EU-funded research project running from 2004 to 2009 with the aim of increasing animal welfare in the food production chain. In this study, OC was compared with the outcome of WQ and AC protocols. To our knowledge, this is the first study to apply different protocols in practice to the same set of farms in Sweden. The aim was to compare

CONTACT B. Staaf Larsson  Birgitta.Staaf.Larsson@slu.se

 Supplemental data for this article can be accessed at <https://doi.org/10.1080/09064702.2024.2317708>

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

how 41 dairy farms were ranked using the three different animal welfare protocols with the starting hypothesis that the best and worse farms in OC also would be identified as the best and worst in AC and WQ, respectively. A second aim was to investigate whether the ranking obtained correlated with data on production, reproduction and mastitis on the farms, with the hypothesis that some correlations with production data would be observed, as earlier reported (Sandgren et al., 2009; Otten et al., 2020). Such information could be used to support implementation of good farm practices. If the AC and WQ protocols indicated poor welfare where OC did not, changes in the Swedish official control could be warranted.

Material and methods

Participating counties and farms

To make the work practically feasible, it was carried out in collaboration with the Swedish CABs, i.e. the regional authorities responsible for the OC (checking animal welfare at farms). At one of their meetings, we had the opportunity to present the project, after which four CABs representing most of Sweden geographically voluntarily agreed to participate in the project. Collaboration with CABs also provided the opportunity to train animal welfare inspectors so that they could implement the WQ protocol. The collaboration with the CABs made it possible to visit the farms and carry out all three

assessments (OC, AC, WQ) at the same time on the same day. The assessments were performed by three assessors in parallel. The study was carried out in spring 2011, before the cows had access to summer pasture. Forty-one farms were recruited from the four CABs. Ten dairy farms in each of the four CAB areas that were already scheduled for an OC inspection were selected for inclusion, plus an additional farm in one county. Accordingly, since the farms already were planned to undergo an animal welfare control by the CABs, farmers did not participate on a voluntary basis primarily, which could induce a bias towards farmers with special interest in animal welfare, and potentially more homogenous farms from an animal welfare perspective. Nevertheless, the farmers agreed to be assessed by the WQ and AC assessment, and all farmers confirmed in writing their willingness to participate. All farms were part of the Swedish National Milk Recording (NMR) system, but had not been assessed previously by the AC or WQ protocol.

Three animal welfare protocols were compared: (1) OC: Swedish official animal welfare control (Swedish Board of Agriculture, Supplementary Material S1); (2) AC: Ask the Cow (Swedish Dairy Association, Supplementary Material S2); and (3) WQ: Welfare Quality (Welfare Quality Network, Supplementary Material S3). An overview of the parameters assessed in these three different animal welfare protocols is provided in Table 1. More information can be found below and in

Table 1. Comparison of the Official Control (OC), Ask the Cow (AC) and Welfare Quality (WC) protocols in terms of animal-based, management-based and resource-based measures included in assessments on farms with dairy cows.

Category	Parameter	OC ¹	AC	WQ ²
Animal-based	Body condition	Yes	Yes*	Yes
	Cleanliness	Yes	Yes*	Yes
	Hooves	Yes	Yes	No
	Hairless patches, and lesions/swellings	No	Yes*	Yes
	Lameness	No	Yes	Yes
	Standing/lying in the stall	No	Yes	No
	Rising behaviour	No	Yes	No
	Lying behaviour	No	No	Yes
	Behaviour	No	To a low extent	Yes
	Vermin, parasites	No	Yes*	Yes
	Health	Partly	Yes*	Yes
	Competition at the feeding table	Yes	Yes	Yes
	Calves, young stock	Yes	Yes*	No
	Avoidance distance	No	No	Yes
	Competent personnel	Yes	No	No
Management-based	Daily observations	Yes	No	No
	Daily/yearly cleaning	Yes	No	No
Resource-based	Water access and hygiene	Yes	Yes	Yes
	Space requirements	Yes	No	No
	The stall interior and floor	Yes	Indirect	Indirect
	Air quality	Yes	No	No
	Day light	Yes	No	No
	Bedding material	Yes	No	No
	Noise	Yes	No	No

¹All questions in the OC checklist are answered 'Yes', 'No', 'Not checked' or 'Not applicable', and relate to all animals on the farm.

²Assessment is done on 35 cows, and in some cases (*) on 35 young stock and 35 calves in addition.

³Number of animals is 30–73, depending on the size of the herd.

Supplementary Material S1–S3. Two of the farms were not included in the WQ assessment, since an avoidance distance test (see below) could not be carried out due to the stall interior (not enough space in front of the cows to approach them properly). Therefore, the WQ assessment was incomplete for those farms. Another farm was excluded from the WQ ranking due to missing data. Since it turned out that the maximum remarks achieved in OC were five, no detailed ranking was feasible, therefore the level of remarks were compared in OC instead of the rankings.

Training of assessors for the study

The OC inspectors and the AC assessors were experienced and worked regularly with the respective assessment program, so no additional training was provided in these cases. Both OC and AC assessors regularly perform calibration exercises. The WQ protocol had not been used in Sweden before this study, so training on the system was provided for experienced OC animal welfare inspectors from the four participating CABs. This comprised a three-day training course given at the animal research facility at the Swedish University of Agricultural Sciences (SLU), Uppsala, and included practical guidance on live animals, exercises involving assessments of animals on video clips and an inter-observer agreement test. The inter-observer agreement test included 28 WQ parameters, performed on live animals (20 cows), regarding body condition score (BCS), cleanliness, health issues, skin condition, lesions and locomotion (Supplementary Table S1). The two instructors, experienced scientists in animal welfare, decided the correct value for each cow and each parameter (silver standard). If the instructors did not agree, the measure was excluded from the calculations. The inter-observer agreement test on the WQ parameters was in agreement with the silver standard to $88 \pm 17\%$ (Range 17–100; Supplementary Table S1).

A total of 24 individuals performed the animal welfare assessments: 10 animal welfare inspectors from the participating CABs carried out OC on farms as routine official inspections, six AC-trained assessors from the Swedish Dairy Association conducted the AC assessments and eight WQ-trained animal welfare inspectors

from the participating CABs conducted the WQ assessments (Table 2). Assessors were only responsible for one protocol each (OC, AC or WQ), and all three assessments of each herd were performed at the same time on the same day. Assessor to be used at the different farms was decided by the regional CAB and for AC by the Swedish Dairy Association.

Animal welfare assessment according to the protocols

Swedish official animal welfare control (OC)

During an OC inspection, animal facilities are checked in terms of e.g. space allowance, lighting, noise and air quality (Supplementary Material S1). In the OC checklists, each parameter corresponds to a statutory requirement, i.e. the animal welfare inspector ticks ‘Yes’ (in compliance with the legislation), ‘No’, ‘Not checked’ or ‘Not applicable’ for the different parameters (Swedish Board of Agriculture, 2022). Questions about water supply to animals in the OC checklist focus on whether the system is designed, dimensioned and positioned so that it allows calm and natural intake, and whether the requirements on access to water and water quality are met (see Supplementary Material S1). The assessment takes around one to two hours, depending on herd size and identified deficiencies.

Ask the Cow (AC)

In the AC protocol, assessments are made at both herd and individual level. At herd level, the Cow Comfort Index (CCI), determined as proportion of cows in cubicles/stalls that are lying down, is assessed, since lying is suggested to be important for good animal welfare (Jensen et al., 2005). Competition between animals for feed is rated as low, medium or high risk of competition behaviour. The number of water bowls or centimetres of water trough is documented, as is the hygiene status of water sources. In addition, the presence of abnormal behaviour, such as stereotypies, is recorded. At the individual cow level, rising behaviour is assessed, where the cow should be able to rise without difficulty with no hesitation longer than five seconds. Body condition score, cleanliness, hoof condition, skin lesions and lameness

Table 2. Assessors (A–Y) applying the official control (OC), Ask the Cow (AC) and the Welfare Quality (WQ) protocols on the 41 participating dairy farms (1–41). Number of farms assessed by each assessor is given in brackets.

Assessment protocol	Farm 1–10	Farm 11–20	Farm 21–31	Farm 32–41
OC	A(5), B(2), C(1), D(1), E(1)	F(10)	G(9), H(2)	I(7), J(3)
AC	K(10),	L(10)	M(11)	N(4), O(4), P(2)
WQ	Q(5), R(5)	S(10) + T(4) ¹	U(11) + V(11) ¹	X(10) + Y(2) ¹

¹The assessments were made in collaboration.

are recorded (see Supplementary Material S2). A farm visit takes around three to four hours.

Welfare Quality (WQ)

The WQ protocol is based on four principles: good feeding, good housing, good health and appropriate behaviour. A WQ assessment starts with an 'avoidance distance test', assessing the human–animal interaction. The assessor approaches cows standing at the feeding table until signs of animal withdrawal emerge or until the assessor can touch the muzzle of the cow (Welfare Quality, 2009; Radeski et al., 2015). If the cow avoids the assessor, the observed distance between assessor and cow is estimated. The assessor also assesses social behaviour (taking into account only aggressive interactions such as head butt, displacement and chasing up), lying behaviour, body condition, cleanliness and lesions. In addition, a qualitative behaviour assessment of how the animals behave and interact with each other is performed by observing the entire herd for 20 min and assessing 20 different parameters on a min–max scale, to gain an overall view of the herd and the expressive quality of the activity at group level. Parameters used at herd level include observations on whether the animals are active, relaxed, fearful, agitated, calm, content, indifferent, frustrated, friendly, bored, playful, positively occupied, lively, inquisitive, irritable, calmless/uneasy, sociable, apathetic, happy or distressed (Welfare Quality, 2009). The number of water bowls or length of water trough is documented (should be more than one water point per cow, according to WQ) and water flow and cleanliness of water sources are noted. For a short version of the WQ assessment protocol, see Supplementary Material S3. The duration of the assessment varies between four and eight hours (Blokhuis et al., 2010), depending on the size of the herd. WQ only assesses dairy cows (adult animals) and contains no protocols for assessing calves and young stock (Brscic et al., 2019).

Herds, production data and health data

The number of cows per herd in the 41 herds ranged from 12 to 268. The largest herd had almost twice as many cows as the second largest ($n = 139$), and the median herd size was 55 cows (mean 65 cows). The average herd size in Sweden at the time was 66 cows (Swedish Board of Agriculture, 2023). Twenty-two of the participating herds had tie-stalls and 19 had a loose-housing system. Yearly milk production was 9554 ± 897 kg energy-corrected milk (ECM) per cow (range 7676–11,855 kg ECM/cow). Cow- and herd-level data were obtained from the NMR database, including

information about milk production, reproduction, health and mastitis history for individual cows.

Data processing and statistical analyses

The AC ranking was calculated from the outcome of 12 parameters: proportion of lean, fat, dirty, severely dirty and lame cows, proportion of cows with long hooves, asymmetric hooves, lesions, severe lesions, rising problems, and proportion of cows lying outside the cubicles and standing in the cubicles, based on the 35 assessed cows. The overall ranking of the herds according to the WQ protocol was based on the sum of scores for the four welfare principles: good feeding, good housing, good health and appropriate behaviour. These calculations were made by staff at the French National Institute for Agriculture, Food, and Environment (INRAE, France), who originally took part in developing the WQ protocol for dairy cows. Based on the outcomes of the WQ and AC assessments, the herds were ranked 1–38 or 41 (where 1 was the best and 38/41 the worst). If two herds received the same value, this was considered a tied ranking and the next number in the series was excluded. The ranking of farms according to OC was based on the number of negative remarks, i.e. less complains resulting in a lower (better) ranking.

The proportions of lean and fat cows were analysed by a *t*-test. The proportions of dirty cows and lesions were not analysed statistically, since the methods for registration differed. Correlation analysis was performed between remarks on OC and rankings in WQ and AC protocols and between remarks/rankings and NMR data (SAS, Version 9.4, SAS Institute Inc., Cary, NC, USA) using the PROC CORR function and regression equations were created in Excel (Microsoft Excel, Microsoft Corp., Redmond, Washington, USA). Rankings of tie-stalls and loose-housing systems according to the different protocols were compared using a Wilcoxon two-sample test (SAS, Version 9.4, SAS Institute Inc., Cary, NC, USA). For correlations and comparisons, *P*-values < 0.05 were considered significant. Results are presented as mean \pm SD if not stated otherwise.

Results

Ranking of herds and correlations between protocols

The number of (negative) remarks per herd in OC ranged from 0 to 5 and the eight farms with most remarks (3–5 remarks) were ranked in the interval 19–40 in AC and 5–37 in WQ (Table 3). There were 18 farms with no remarks

Table 3. Number of negative remarks made about the 41 dairy farms in the Official Control (OC) and farm ranking according to the Ask the Cow (AC) and Welfare Quality (WQ) protocols. There are three missing values for WQ, two since avoidance distance not could be assessed on two farms and one due to missing values, so the total value could not be calculated.

Remarks in OC	Rank AC	Rank WQ
0	1	16
0	6	35
0	8	23
0	9	17
0	11	29
0	12	21
0	13	3
0	17	
0	17	6
0	19	14
0	23	18
0	24	20
0	26	12
0	28	36
0	30	33
0	32	8
0	33	1
0	36	
1	1	7
1	5	25
1	7	13
1	15	2
1	16	38
1	19	28
1	25	31
1	29	4
1	34	15
2	3	10
2	3	9
2	9	30
2	14	34
2	31	11
2	41	32
3	19	22
3	27	5
4	19	27
4	40	26
5	35	19
5	37	
5	38	24
5	39	37

in OC, and these were within the ranking range 1–36 in both AC and WQ (Table 3). There was a weak to moderate positive correlation ($r = 0.40$, $P = 0.009$) between remarks in OC and the AC ranking, but no correlation ($P > 0.05$) between OC remarks and the WQ ranking, or between the AC and WQ rankings (Figure 1).

The 41 herds as a whole received 55 remarks in OC out of 1763 possible (41 farms \times 44 parameters, i.e. 3%), the most common ($n = 11$) being lack of a back-up system to ensure sufficient air regeneration in a mechanically ventilated animal house and/or lack of an alarm system to signal a ventilation breakdown. The second most common remark ($n = 8$) concerned calf housing, e.g. calves were kept in individual pens after eight weeks of age, which is not permitted under EU legislation (Council Directive 2008/119/EC of 18

December 2008, European Commission, 2008), and thus not under Swedish legislation. In AC, the 41 herds received 624 remarks out of 1876 possible (33%), with the most common remarks concerning lesions (39 of 41 farms, mean 28% of cows) and asymmetrical hooves (33 farms, mean 23% of cows). In WQ, there were 680 remarks out of 1083 possible (63%), the most frequent being that all 41 farms had cows with dirty legs (mean 68% of cows at herd level), dirty hindquarters (50%), hairless patches (61%) or lesions (32%).

Tie-stalls versus loose-housing system

In WQ, tie-stalls were rated significantly lower (mean ranking 25 ± 10) than loose-housing systems (mean ranking 13 ± 10) ($P = 0.001$), but there were no differences according to AC ($P = 0.5$, mean ranking 20 ± 13 for tie-stalls and 22 ± 11 for loose-housing systems) or OC ($P = 0.7$, mean number of negative remarks 1.4 ± 1.7 for tie-stalls and 1.3 ± 1.7 for loose-housing systems).

Correlations between protocols and NRM data

No correlations were found between the NRM data (kg ECM, calf mortality, mastitis, percentage of cows with >70 days between calving and first insemination) and the ranking obtained with the different protocols, with the exception of percentage of cows failing to become pregnant within 120 days of calving, which was correlated with the OC remarks ($r = 0.36$, $P = 0.02$; Supplementary Table S2). See Supplementary Table S3 for all mean values (\pm SD) of the three protocols and the NRM data.

Comparison between protocols on body condition

There were no remarks about lean or fat cows for any of the 41 farms when assessed with OC, while there were 18 farms with no lean cows according to AC and eight according to WQ (Figure 2). The 23 farms with lean cows according to AC had on average 6% lean cows (range 2–17%). The corresponding figure for WQ was 33 farms with on average 8% lean cows (range 2–56%). There were no differences in the proportion of lean cows in AC and WQ ($3 \pm 4\%$ and $7 \pm 10\%$, respectively; $P = 0.5$). There was a significant difference between cows scored as fat by AC and WQ ($6 \pm 8\%$ and $4 \pm 5\%$, respectively; $P = 0.02$). There were 15 farms with no fat cows according to AC and 18 according to WQ (Figure 2). The 26 farms with fat cows according to AC had on average 9% fat cows (range 3–40%). The corresponding figure for WQ was 23 farms with on average 7% fat cows (range 1–15%).

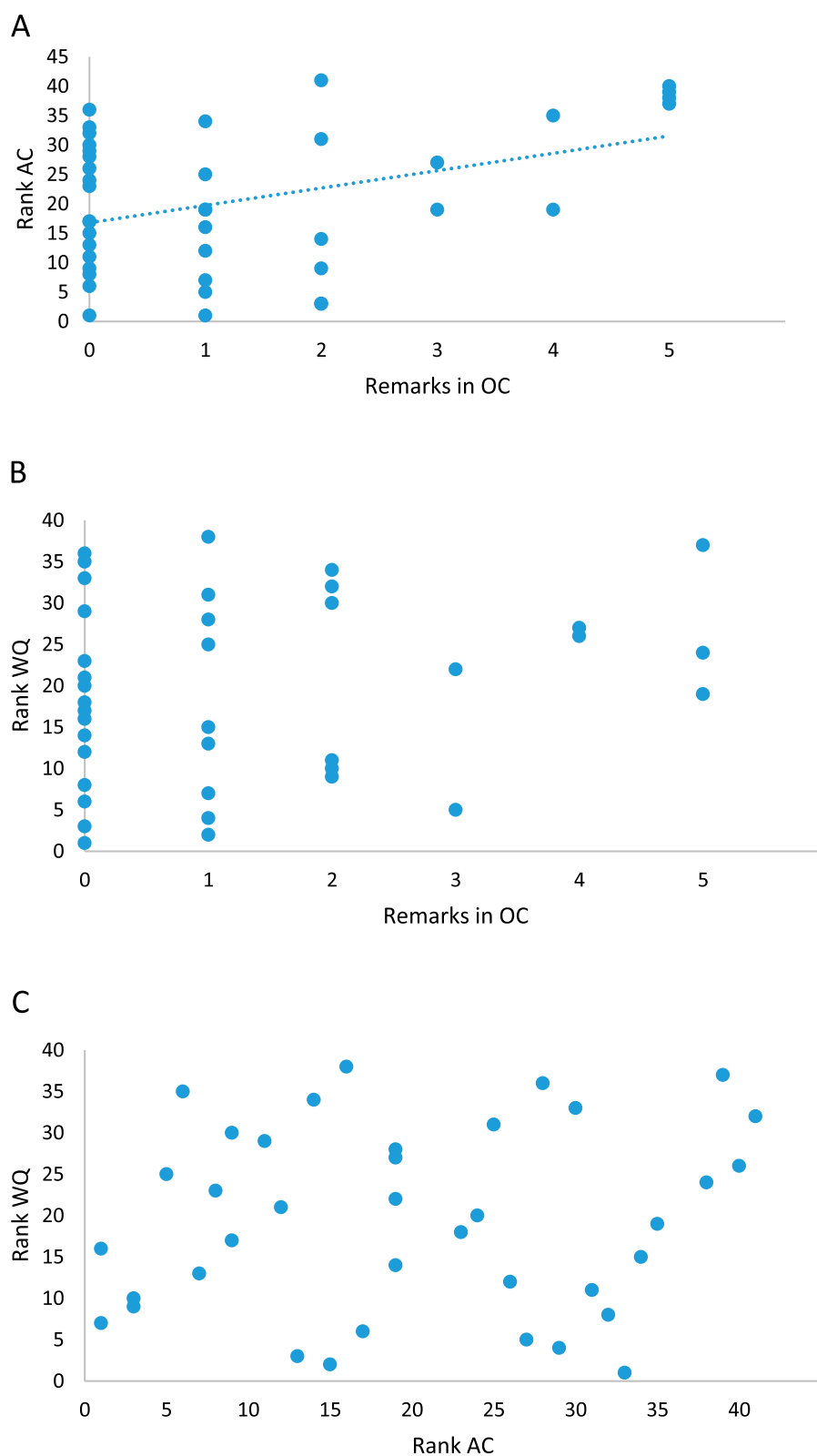


Figure 1. Correlation between (A) number of remarks in the Official Control (OC) and ranking of the 41 dairy farms using the 'Ask the Cow' (AC) protocol ($r = 0.40$, $P = 0.009$, $y = 2952x + 16,796$), (B) number of remarks in OC and ranking using the Welfare Quality (WQ) protocol ($P > 0.05$) and (C) the AC and WQ rankings. Correlations based on 41 farms, except for WQ ($n = 38$ farms).

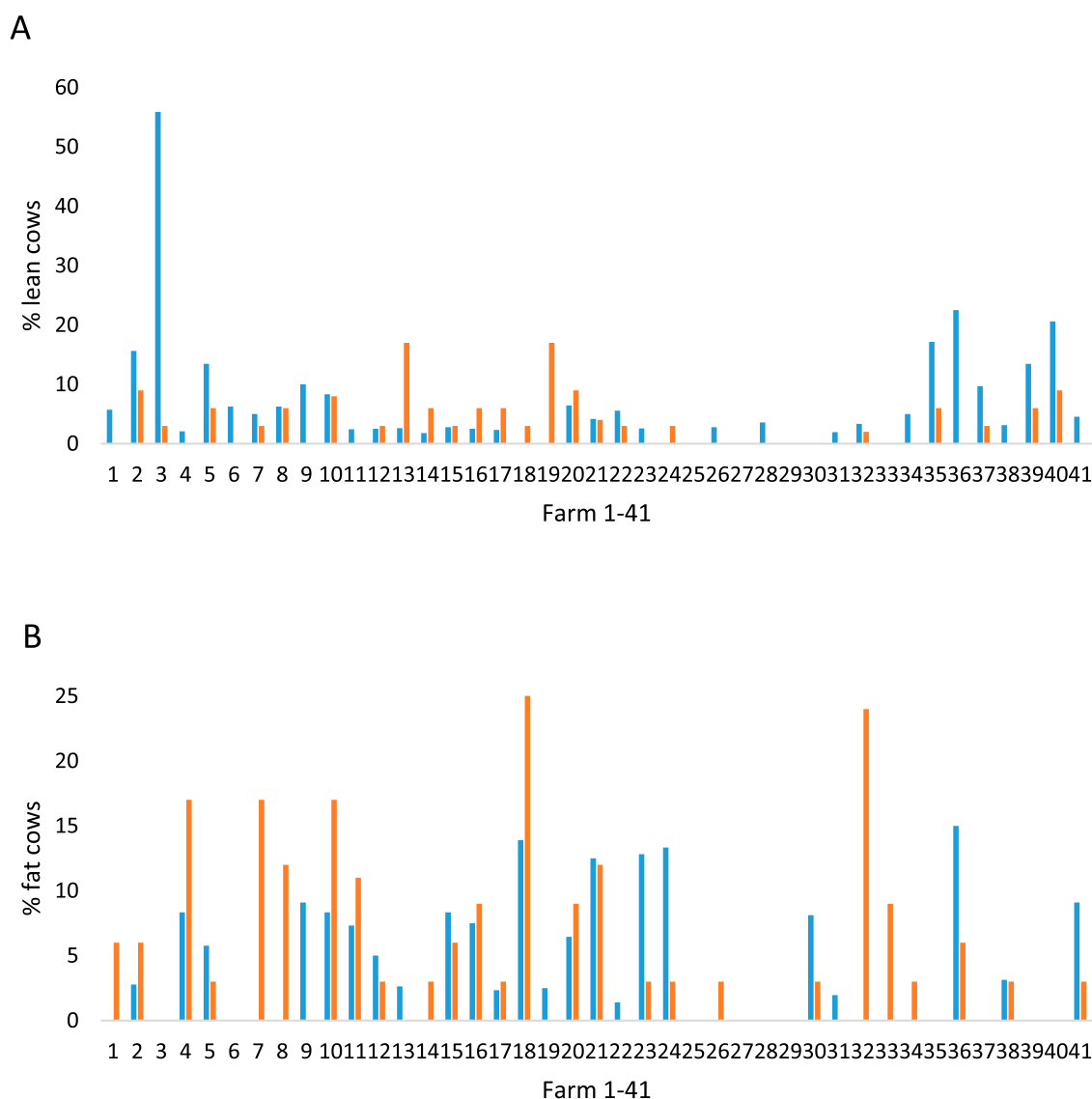


Figure 2. Proportion of lean cows (A) and fat cows (B) per herd in the Ask the Cow (AC) assessment (red bars) and Welfare Quality (WQ) assessment (blue bars). Farm number 1–41 is not related to the ranking.

Comparison between protocols on cleanliness

Only two herds received remarks about dirty cows in the OC assessment, but according to the AC assessment 31 herds out of 41 had dirty cows. In two herds, almost all cows (97%) were assessed as dirty in AC, and cows were assessed as severely dirty in seven herds. Ten herds did not have any dirty animals at all, according to the AC assessment (Figure 3). The 31 herds with dirty cows according to AC had on average 23% dirty cows (range 3–97%) and seven herds had severely dirty cows (mean 7%, range 3–20%).

According to the WQ assessments, all herds had cows with dirty legs and flanks (mean 68% and 50% respectively, range 8–100% and 4–97%, respectively), and all herds except one had cows with dirty udders (Figure 3).

The 40 herds with dirty udders according to WQ had on average 44% cows with dirty udders (range 3–85%).

Comparison between protocols on skin lesions

The proportion of cows assessed as having skin lesions varied between 2% and 75% ($32 \pm 17\%$) in WQ and 0 and 68% ($28 \pm 17\%$) in AC (Figure 4). The proportion of cows with severe lesions according to AC was $5 \pm 7\%$ and varied between 0% and 26%. Cows on two farms had no lesions and cows on 16 farms had no severe lesions according to AC. The 39 herds with lesions had 6–68% (mean 30%) cows with lesions, and the 25 herds with severe lesions according to AC had 3–26% (mean 13%) cows with severe lesions. In WQ

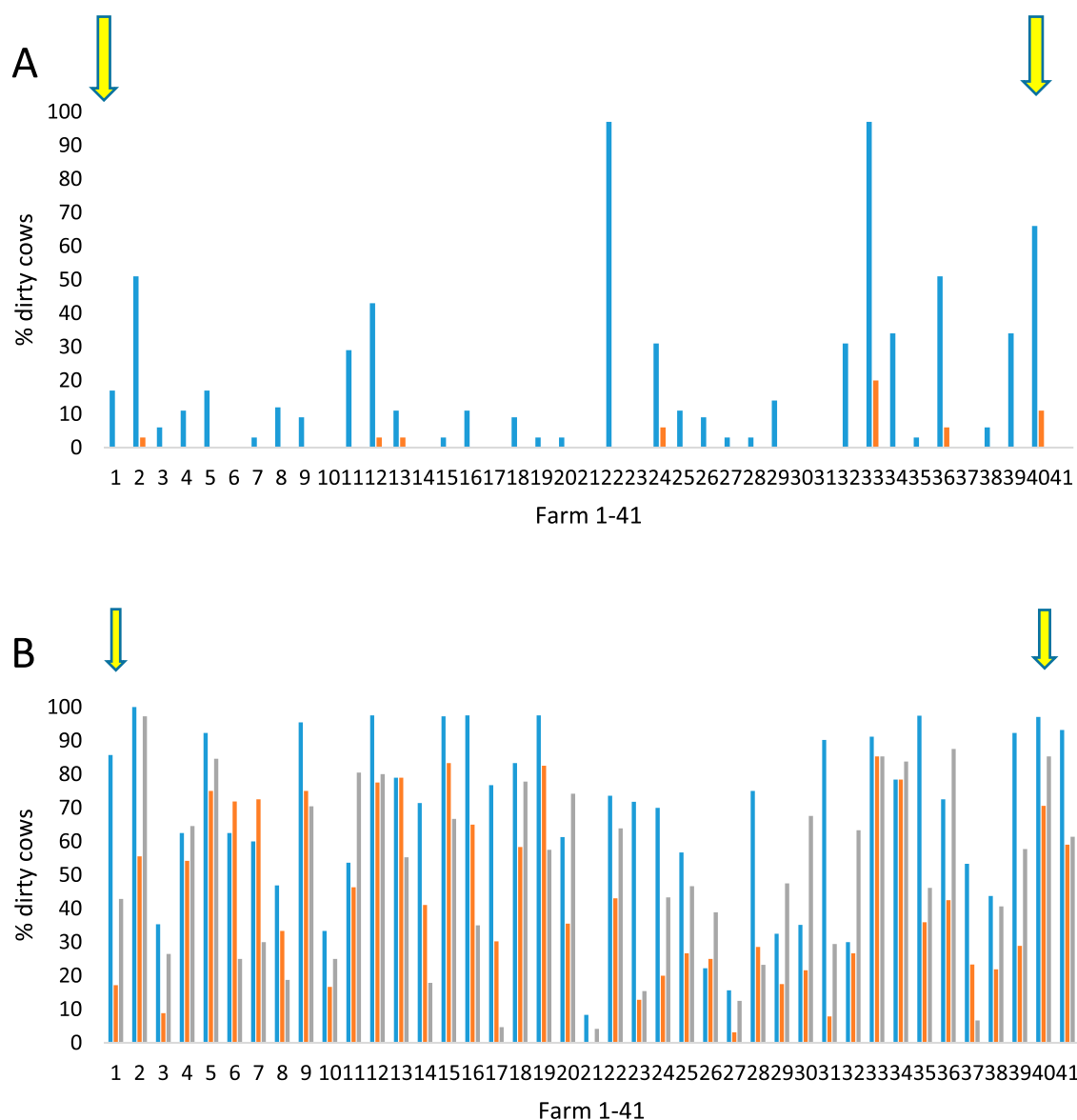


Figure 3. Proportion of (A) dirty cows (blue bars) and severely dirty cows (red bars) per herd in the Ask the Cow (AC) assessment and (B) proportion of cows per herd with dirty legs (blue bars), dirty udder (red bars) and dirty hindquarters (green bars) in the Welfare Quality (WQ) assessment. Farms that received negative remarks in Official Control (OC) are indicated with yellow arrows. Farm number 1–41 is not related to the ranking.

assessments, no farm was found to have any lesions in the herd. There was no correlation ($P > 0.05$) regarding lesions between AC and WQ. Skin lesions are not covered by the OC checklist.

Water supply and water quality

In OC, only one farm received a remark regarding water supply and/or the quality of the water. In AC assessments, seven farms received remarks on water hygiene and nine did not have a sufficient supply of water according to the AC protocol. In WQ assessments, three of the farms were considered not to have

sufficient cleanliness of the water source and 31 farms were categorised as not having sufficient water flow.

Discussion

The aim of this study was to compare how 41 dairy farms were ranked using three animal welfare protocols, and to our knowledge, this is the first study in Sweden to do this. The starting hypothesis in this study, that the same farms would be identified as best and worse irrespective of the protocol used, was not supported by the results. The ranking of the farms differed substantially between the assessment systems, although a

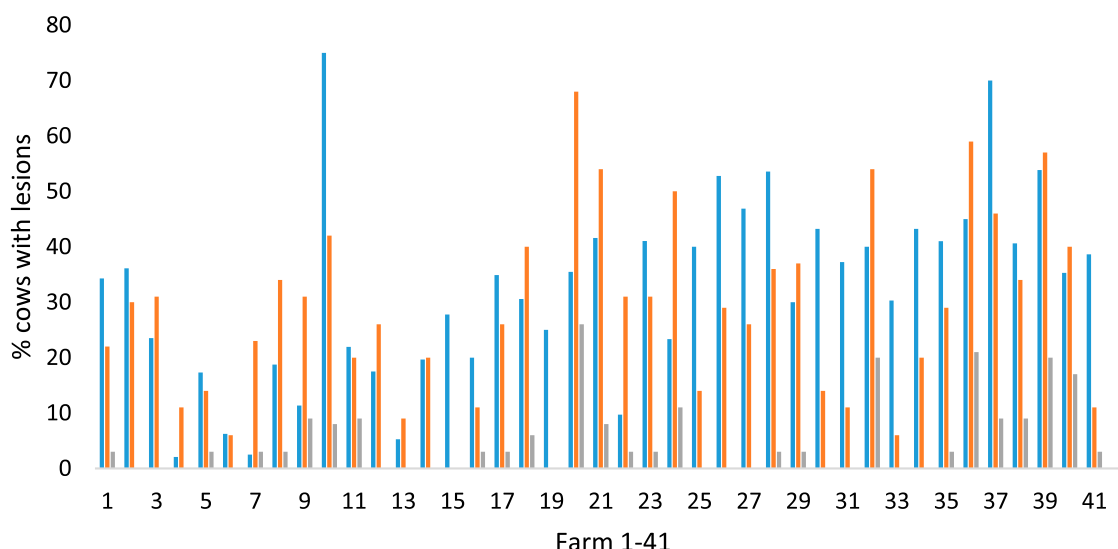


Figure 4. Proportion of cows with skin lesions in the Welfare Quality (WQ) assessment (blue bars) and Ask the Cow (AC) assessment (red bars = skin lesions, green bars = severe skin lesions). Farm number 1–41 is not related to the ranking.

moderate positive correlation was found between the outcomes of AC and OC. In contrast to our study, Otten et al. (2020) reported high agreement between two protocols, but both these protocols mainly used similar measurements (i.e. WQ or WQ based). The low agreement between the outcomes of the three protocols in our study is probably an effect of the use of different parameters, different scoring systems and scales, and different sample sizes. Stull et al. (2005) compared three assessment programs on 10 dairy farms in California, USA. The three assessment programs did not rank the farms similar, although all programs identified the same two farms with the lowest ranking.

However, the use of different assessors for the three protocols, might also have had impact on the agreement. A previous study by Knierim & Winckler (2009) raised the challenge of low reliability when assessing animal-based welfare parameters on farms and pointed out that robust agreement over time and/or between assessors is difficult to achieve (except for the ‘avoidance distance’ parameter in their study). However, the WQ inter-observer test in this study showed relatively high agreement between assessors. Another methodological strategy could have been to cross-train all observers on all three programs and randomly assign them to use a particular assessment type on a given farm. However, this was not practically feasible in the present study because the 41 farms were spread across Sweden. A strength of the present study is that experienced professional assessors carried out the assessments on the same day on each farm, which ensured professionalism and that conditions were the same although different protocols were used. As

explained earlier, no joint training were performed within OC and AC, since their inspectors were experienced professional assessors. However, an inter-observer test would have been beneficial to investigate whether there were any differences in assessment outcomes between assessors.

In the OC system, the assessment of animal welfare is based to a low extent on animal-related parameters (e.g. cleanliness and body condition), with no documentation in the assessment protocol on number of animals affected and level of body condition. In OC, there were no remarks on body condition, whereas the AC and WQ outcomes showed that several cows were considered both lean and fat. There was no significant difference in the proportion of lean cows between the AC and WQ protocols (3% lean cows in AC and 7% in WQ), indicating that the two protocols work similarly for this parameter. The proportion of lean cows was similar to that in a previous study using WQ (Tremetsberger et al., 2019). In our study, one farm in the WQ protocol deviated by more than 3 SD from the mean regarding lean cows. However, that was the first WQ assessment by the evaluator and unfamiliarity with the protocol or the recording system might have influenced the rating. Another farm deviated by more than 3 SD from the mean regarding fat cows in the AC protocol, with 40% of the cows recorded as fat. However, in that case the assessor also made a comment in text about ‘well-fed cows’ and it can therefore be assumed that the high number of fat cows was according to that evaluator’s opinion, while only 14% of the cows on the same farm were recorded as fat according to the WQ protocol/evaluator.

In both AC and WQ, substantial criticisms about cleanliness were made, i.e. in WQ all 41 farms had cows with dirty legs and dirty hindquarters and in AC 31 herds out of 41 had dirty cows. In OC, only two herds received comments about dirty cows. Both AC and WQ thoroughly evaluate cleanliness and the difference in outcome at herd level in this study was probably because the assessment criteria are different and the sample of animals within a herd was based on different individuals. It may also have been partly due to the dissimilar systems, with 12–35 cows assessed in AC and 12–73 in WQ, and to different assessors evaluating the cows differently. When assessing cleanliness, the entire body is checked in AC, whereas only one side of the body is checked in WQ. A study on Norwegian farms by Hauge et al. (2012) concluded that good housing, feeding and management are essential for keeping animals clean. In Austrian dairy herds, Tremetsberger et al. (2015) found that udder health and cleanliness improved significantly when implementing better husbandry practices, such as enhanced udder cleaning routines.

Lesions were registered in WQ and AC but there were no correlations regarding lesions between the two protocols. This was possibly due to different animals being measured or to the measuring procedure used. As previously mentioned, skin lesions, wounds and lameness are not included in the OC checklist. One lesion may occur by chance, but lesions on multiple cows suggest the presence of a systemic problem potentially leading to pain. By mainly focusing on resource-based measures and farmers' compliance with the legislation, OC excludes much information about individual animal welfare. In the OC checklist, the only question concerning sick animals or animals with lesions is whether the animals are given the necessary care and have access to a separate compartment. The checklist contains no items about the presence or proportion of sick animals or animals with lesions. Therefore, we suggest that including animal-based measurements of skin lesions, wounds and lameness could improve OC and the welfare of individual cows. No conclusions could be reached regarding the protocol with the best validity as regards lesions, so further studies on methodology are needed to determine which measures should be included in OC.

The most common remarks in OC were not animal-based but concerned with building ventilation. This is in agreement with the outcome of the official control in Sweden during the study year, where remarks about the ventilation system were the most common type of negative observation (K. Andersson, Swedish Board of Agriculture, pers. Comm. 17 January 2021). In contrast, neither AC nor WQ registered anything with respect to ventilation of the building, stable temperature, air

quality or even animal-based indicators of poor ventilation or thermal comfort. We suggest a combination of resource and animal-based registrations to better monitor thermal comfort and air quality.

All three assessment systems studied include a resource-based measurement of water supply. As mentioned, water supply and water quality are checked in OC, while in AC and WQ the number of water bowls or length of water troughs available is recorded, as is the hygiene status of the water source. In addition, WQ measures water flow. The WQ system often gives tie-stalls a low ranking for the principle of good nutrition, because tied cows often have access to only one water bowl. In a study comparing WQ and the Danish Animal Welfare Index (DAWIN), Otten et al. (2020) concluded that insufficient water supply was the main area of concern according to the WQ protocol. On studying 92 dairy farms in England and Wales, Heath et al. (2014) found that the one resource-based parameter (absence from prolonged thirst) measured in WQ correctly classified 88% of the farms in the same way as the whole WQ assessment.

Tie-stalls were ranked worse than loose-housing systems with the WQ protocol, but no difference in ranking was observed with the AC or OC protocols. The reason for the poor ranking in WQ was that tied cows often had access to only one water bowl and were less able to express natural behaviours, since they had no access to daily exercise and were only kept loose when grazing during summer. Only one of the 22 herds with tie-stalls in this study offered the cows exercise all year around.

An animal welfare protocol used for official control must be based on validated methods to register animal welfare, have legal certainty but also be practically feasible to perform at the farm level. One challenge with the WQ assessment is that the behaviour tests are time-consuming (de Vries et al., 2013). Both the avoidance test evaluating the human–animal interaction, standing still looking at the animals' behaviour during the quality behaviour observation (20 min), and waiting for the cows for measuring and evaluating the lying down behaviour takes time. All assessors in our study reported that WQ requires too much time to be implemented fully in OC. Inclusion of new registrations to a protocol (e.g. OC) must therefore be made by weighing time consumption and the value and quality of the registrations for assessing animal welfare.

Conclusions

There were marked differences in assessment outcomes for the three protocols compared in this study, which is

unsatisfactory from a legal perspective. Contrary to our hypothesis that the worst (with most remarks) and best (with fewest remarks) farms would be identified similarly by all three protocols, the results obtained showed the opposite, although there was a moderate correlation between the rankings in AC and OC. Substantial remarks on body condition and cleanliness were made according to AC and WQ, but not according to the OC system, with which remarks concerning the robustness of on-farm ventilation systems were instead most common. Skin lesions, wounds and lameness, which are not even measured in OC, received high numbers of remarks according to AC and WQ. We therefore suggest including animal-based measurements of skin lesions, wounds and lameness in order to improve OC and the welfare of individual cows. Including the amount of animals affected, and the level of dirtiness and body condition in the assessment would further improve the OC protocol. However, the outcome of animal-based assessments is highly dependent on the methods used and further studies are needed to develop and validate methods.

Ethics approval

Ethical review and approval from an animal ethics committee were not required for this study, since Swedish regulations explicitly state that animal experiments using privately owned animals kept in their normal environment and not subjected to any painful or stressful procedures may be conducted without an operating licence or ethical approval for animal research.

Data and model availability statement

The datasets considered used in this study contain sensitive information and cannot be made publicly accessible, but they are available (with participating farms anonymised) from the corresponding author upon reasonable request.

Acknowledgements

We are grateful to the farmers who participated in this study and to the animal welfare inspectors and all other participants for making the study possible. Thanks also to Christoph Winckler and Lukas Tremetsberger, University of Natural Resources and Life Sciences (BOKU) Austria, who performed the WQ training; Isabelle Veissier and Anne Lamadon, French National Institute for Agriculture, Food, and Environment (INRAE) France, who conducted the WQ calculations; Claudia von Brömssen at the Department of Energy and Technology, Swedish University of Agricultural Sciences, for statistical advice; and Mary McAfee for English editing. Part of this study has previously been published as the MSc thesis of the first author (Staaf Larsson, 2014).

Disclosure statement

Louise Winblad von Walter is employed at Växa Sverige, former Swedish Dairy Association, but she was not involved in the collection of data or analysis of the data. Birgitta Staaf Larsson, Anna Jansson, Mia Holmberg, Margareta Stéen and Kristina Dahlborn have no competing interests to declare.

Funding

This work was supported by the Swedish Government [grant number 1:23 anslagspost 7, punkt 13]; Svenska Djurskyddsföreringen [grant number 2014-05-06].

Financial support statement

This work was supported by the Swedish Government through the Swedish Centre for Animal Welfare, the Swedish University for Agricultural Sciences, and by the Swedish Association for the Protection of Animals (for preparation of this article).

ORCID

B. Staaf Larsson  <http://orcid.org/0000-0002-6748-6047>
A. Jansson  <http://orcid.org/0000-0001-5612-9020>
M. Holmberg  <http://orcid.org/0000-0002-2497-7647>
L. Winblad von Walter  <http://orcid.org/0000-0002-7966-1842>
M. Stéen  <http://orcid.org/0000-0001-8224-0224>
K. Dahlborn  <http://orcid.org/0000-0002-1029-9013>

References

- Barry, C., Ellingsen-Dalskau, K., Garmo, R. T., Grønmo Kischel, S., Winckler, C. & Kielland, C. (2023). Obtaining an animal welfare status in Norwegian dairy herds – A mountain to climb. *Frontiers in Veterinary Science*, 10, 1125860. doi:10.3389/fvets.2023.1125860
- Berthe, F., Vannier, P., Have, P., Serratos, J., Bastino, E., Broom, D. M., Hartung, J. & Sharp, J. M. (2012). Special issue: The role of EFSA in assessing and promoting animal health and welfare. *EFSA Journal*, 10(10), s1002. (10 pp.). doi:10.2903/j.efsa.2012.s1002. Available online: www.efsa.europa.eu/efsajournal.
- Blokhuys, H. J., Veissier, I., Miele, M. & Jones, B. (2010). The welfare quality® project and beyond: safeguarding farm animal well-being. *Acta Agriculturae Scandinavica, Section A-Animal Science*, 60(3), 129–140.
- Brscic, M., Otten, N. D., Contiero, B. & Kirchner, M. K. (2019). Investigation of a standardized qualitative behaviour assessment and exploration of potential influencing factors on the emotional state of dairy calves. *Animals*, 9(10), 757.
- de Vries, M., Engel, B., den Uijl, I., van Schaik, G., Dijkstra, T., de Boer, I. J. M. & Bokkers, E. A. M. (2013). Assessment time of the Welfare Quality (protocol for dairy cattle). *Animal Welfare*, 22, 85–93.
- European Commission. (1998). Council Directive 98/58/EC of 20 July 1998 concerning the protection of animals kept for farming purposes. *Official Journal of the European Communities*, L 221, 8.8.98, 23–27.

- European Commission. (2008). Council Directive 2008/119/EC of 18 December 2008 laying down minimum standards for the protection of calves. *Official Journal of the European Communities*, L 10, 15.1.2009, 8.
- Gieseke, D., Lambertz, C. & Gauly, M. (2022). Effects of housing and management factors on selected indicators of the welfare quality[®] protocol in loose-housed dairy cows. *Veterinary Sciences*, 9, 353. doi:10.3390/vetsci9070353
- Hauge, S. J., Kielland, C., Ringdal, G., Skjerve, E. & Nafstad, O. (2012). Factors associated with cattle cleanliness on Norwegian dairy farms. *Journal of Dairy Science*, 7, 3213–3282.
- Heath, C. A. E., Browne, W. J., Mullan, S. & Main, D. C. J. (2014). Navigating the iceberg: reducing the number of parameters within the Welfare Quality[®] assessment protocol for dairy cows. *Animal*, 8(12), 1978–1986.
- Jensen, M. B., Pedersen, L. J. & Munksgaard, L. (2005). The effect of reward duration on demand functions for rest in dairy heifers and lying requirements as measured by demand functions. *Applied Animal Behaviour Science*, 90, 207–217. doi:10.1016/j.applanim.2004.08.006
- Keeling, L. J. (2009). An analysis of animal-based versus resource-based comments in official animal welfare inspection reports from organic and conventional farms in Sweden. *Animal Welfare*, 18, 391–397.
- KilBride, A. L., Mason, S. A., Honeyman, P. C., Pritchard, D. G., Hepple, S. & Green, L. E. (2012). Associations between membership of farm assurance and organic certification schemes and compliance with animal welfare legislation. *Veterinary Record*, 170, 152.
- Knierim, U. & Winckler, C. (2009). On-farm welfare assessment in cattle: validity, reliability and feasibility issues and future perspectives with special regard to the Welfare Quality[®] approach. *Animal Welfare*, 18(4), 451–458.
- Lundmark, F., Röcklinsberg, H., Wahlberg, B. & Berg, C. (2016). Content and structure of Swedish animal welfare legislation and private standards for dairy cattle. *Acta Agriculturae Scandinavica Section A-Animal Science*, 66(1), 35–42.
- Otten, N. D., Rousing, T., de Oliveira, V. H. S., Reiten, M., Michelsen, A. M., Hakansson, F., Lund, V. P., Houe, H., Denwood, M., Sorensen, J. T., Forkman, B. & Kirchner, M. K. (2020). A weak point analysis of welfare in Danish dairy herds using two different welfare assessment systems. *Animal Welfare*, 29(2), 197–207.
- Radeski, M., Janevski, A. & Ilieski, V. (2015). Screening of selected indicators of dairy cattle welfare in Macedonia. *Macedonian Veterinary Review*, 38(1), 43–51.
- Sandgren, C. H., Lindberg, A. & Keeling, L. J. (2009). Using a national dairy database to identify herds with poor welfare. *Animal Welfare*, 18, 523–532.
- Staaf Larsson, B. (2014). *Comparison of three assessment programmes in Sweden; The official animal-based welfare control, Ask the Cow and Welfare Quality[®]*. MSc thesis, Gothenburg, Sweden: University of Gothenburg.
- Stull, C. L., Reed, B. A. & Berry, S. L. (2005). A comparison of three animal welfare assessment programs on California dairies. *Journal of Dairy Science*, 88(4), 1595–1600. doi: 10.3168/jds.S0022-0302(05)72828-9. PMID: 15778329.
- Swedish Board of Agriculture. (2019). Statens jordbruksverks föreskrifter och allmänna råd (SJVFS 2019:18) om nötkreaturshållning inom lantbruket m.m. [The Swedish Board of Agriculture's regulations and general recommendations (SJVFS 2019:18) on cattle husbandry in agriculture, etc.], Case No L 104.
- Swedish Board of Agriculture. (2022). *Vägledning för offentlig djurskyddskontroll* ['Guidelines for official animal welfare control']; Swedish Board of Agriculture: Jönköping (Sweden).
- Swedish Board of Agriculture. (2023). Antal djur och jordbruksföretag med djur efter Län, Kategori, Variabel, Tabelluppgift och År. PxWeb (sjv.se). [Number of animals and agricultural enterprises with animals by County, Category, Variable, Tabular information and Year.] Assessed 1 February 2023.
- Swedish Government. (2018). Animal Welfare Act (2018:1192).
- Swedish Government. (2019). Animal Welfare Ordinance (2019:66).
- Tremetsberger, L., Leeb, C. & Winckler, C. (2015). Animal health and welfare planning improves udder health and cleanliness but not leg health in Austrian dairy herds. *Journal of Dairy Science*, 98(10), 6801–6811.
- Tremetsberger, L., Winckler, C. & Kantelhardt, J. (2019). Animal health and welfare state and technical efficiency of dairy farms: possible synergies. *Animal Welfare*, 28(3), 345–352.
- Van Os, J. M. C., Winckler, C., Trieb, J., Matarazzo, S. V., Lehenbauer, T. W., Champagne, J. D. & Tucker, C. B. (2018). Reliability of sampling strategies for measuring dairy cattle welfare on commercial farms. *Journal of Dairy Science*, 101, 1495–1504. doi:10.3168/jds.2017-13611
- Welfare Quality[®]. (2009). *Welfare quality[®] assessment protocol for cattle. welfare quality[®]*. Consortium, Lelystad, Netherlands. http://www.welfarequality.net/media/1088/cattle_protocol_without_veal_calves.pdf.