© 2020 Universities Federation for Animal Welfare
The Old School, Brewhouse Hill, Wheathampstead,
Hertfordshire AL4 8AN, UK
www.ufaw.org.uk

Animal Welfare 2020, 29: 197-207
ISSN 0962-7286
doi: 10.7120/09627286.29.2.197

A weak point analysis of welfare in Danish dairy herds using two different welfare assessment systems

ND Otten*[†], T Rousing[‡], VHS de Oliveira[‡§], M Reiten[‡], A-M Michelsen[†], F Hakansson[†], VP Lund[†#], H Houe[†], M Denwood[†], JT Sørensen[‡], B Forkman[†] and MK Kirchner[†¶]

[†] Department of Veterinary and Animal Sciences, University of Copenhagen, Denmark
[‡] Department of Animal Science, Aarhus University, Blichers Alle 20, 8830 Tjele, Denmark
[§] National Veterinary Institute, Department of Disease Control and Epidemiology, SE-751 89, Uppsala 7, Sweden
[#] The Danish Veterinary Association, Peter Bangs Vej 30, 2000 Frederiksberg, Denmark
[¶] VIER PFOTEN, Stiftung für Tierschutz, Linke Wienzeile 236, A-1150, Vienna, Austria
* Contact for correspondence: nio@sund.ku.dk

Abstract

This study aimed to identify current weak points in animal welfare in Danish dairy production at herd level using the Welfare Quality® (WQ) protocol, and at national level using the Danish Animal Welfare Index (DAWIN) protocol. The DAWIN was developed as a monitoring tool for the welfare of the Danish dairy cow population, derived from the aggregation of DAWIN assessments at herd level. The DAWIN dairy cow protocol covers 29 measures (13 resource- and 16 animal-based measures) that were weighted and aggregated into a final overall population welfare score. A total of 3,591 cows from 60 dairy herds were assessed throughout 2015. Results from both the WQ and DAWIN were presented at six criteria levels in order to identify specific areas of concern relating to animal welfare at herd versus population level. Both protocols indicated a good general level of welfare across study herds, but also identified insufficient water supply as the main area of concern. In addition, resting comfort (ie time needed to lie down, collisions with barn equipment, cleanliness of rear body parts, animals lying outside of the designated lying area) and disease (in terms of the proportion of cows with chronically elevated somatic cell counts) were identified as problematic areas. The two assessment protocols both identified behavioural deficits, but in the WQ it was due to zero-grazing systems in contrast to the insufficient numbers of cow brushes in the DAWIN protocol. Despite differences in the aggregation, similar areas of concern were identified at criteria level.

Keywords: *animal welfare, assessment, dairy cattle, population level, welfare problems, Welfare Quality®*

Introduction

Animal welfare assessment is a complex task, calling for scientifically valid multi-criteria and multi-stage approaches. The complexity starts with the definition of animal welfare, as this is often thought to encompass several important dimensions (Fraser *et al* 1997; Appleby & Sandøe 2002). While the underlying definition of animal welfare will determine the design of the assessment, the intended purpose must also be reflected. Motives for assessing animal welfare vary considerably, as they can cover anything from classification and certification to decision support. The Animal Needs Index (ANI; Bartussek 1999), for instance, is performed at a federal level in Austria as part of an organic compliance control, while the RSPCA's Freedom Food in the UK is intended as a product labelling system. Finally, animal welfare assessment may be intended as an advisory tool for producers, similar to the Danish Cattle Federations (DCF) protocol (Danish Cattle Federation 2005). The choice of measures and aggregation methods used in the given welfare assessment protocols are highly dependent on the intended purpose of these protocols (Johnsen & Sandøe 1999). Some protocols therefore rely more heavily than others on the assessment of risk factors for impaired animal welfare in terms of resource- and management-based measures. The resource-based measures are valuable in decision support schemes, as these risk factors can be altered to achieve better results in selected animal-based measures, and they are less time consuming to obtain. However, the scientific consensus is that a truer picture of animal welfare can be achieved by observing the unit of interest, ie observing the animals directly by means of animal-based measures (Webster *et al* 2004; Keeling 2009). This was the approach taken in the most comprehensive welfare assessment protocol to date, the Welfare Quality® (WQ) protocol. The WQ project developed welfare assessment protocols for several animal species within the primary production (Blokhuis 2008), and the protocols relied primarily on animal-based measures for evaluating animal welfare at herd level.

While most protocols aim to evaluate animal welfare at herd level, to our knowledge no previous studies have evaluated the welfare of a whole animal population, eg at a national

https://doi.org/10.7120/09627286.29.2.197 Published online by Cambridge University Press

level. National animal welfare monitoring traditionally consists of state veterinarians conducting official animal welfare audits. In Denmark, these are performed as direct assessments of compliance with legislation related to animal welfare, ie compliance with the minimum standards for housing and keeping the given species. Controls can be based on resource and management parameters or as focused control campaigns, eg audits of compliance with legislation on keeping sick and injured dairy cows, or the roadside control of animal transport (Danish Veterinary and Food Administration 2018). In order to complement official controls and enable monitoring of changes in animal welfare for pigs and dairy cattle at a national level, the University of Copenhagen (KU) and Aarhus University (AU), together with the Danish Veterinary and Food Administration (DVFA), have developed national animal welfare indices for six different groups of production animals, known as the Danish Animal Welfare Index (DAWIN). Using this novel approach, authorities intend to embrace a more complete monitoring of animal welfare, rather than solely regulating legislative compliance.

The overall aim of DAWIN was to establish welfare indices in order to improve animal welfare for production animals by using the DAWIN to identify and specifically target problem areas in the primary production. Hence, the present paper is the first to describe the DAWIN approach for dairy cattle. During the period 2013–2016, KU and AU worked closely together on the development of six indices for: i) dairy cows; ii) dairy calves; iii) lactating sows; iv) piglets; v) weaners and finisher pigs; and vi) gestating sows and gilts. For more detailed information on the pig indices, see Michelsen *et al* (submitted). It was decided that the DAWIN protocol would be based on a hedonistic definition of welfare in which affective states define animal welfare. Just as in the WQ, the DAWIN also centres on the experiences of the animal but is specifically focused on Danish production settings. For operational reasons, the dairy cow protocol was a simplified version of the WQ protocol, as certain criteria were either not feasible or not applicable within the scope of the assessment due to it being at national level in contrast to the herd-level approach used within the WQ. The DAWIN for dairy cows therefore represents the following six welfare criteria from the WQ: 'Absence of prolonged hunger'; 'Absence of prolonged thirst'; 'Comfort around resting'; 'Ease of movement'; 'Absence of disease'; and 'Expression of other behaviours'. A literature review initially identified 79 measures related to dairy cow welfare and concerning the 12 overarching WQ criteria. A final list of 29 measures was created after evaluation in terms of validity, repeatability, feasibility and coverage of multiple criteria. A total of 13 resource- and 16 animal-based measures were included in the evaluation of the six welfare criteria. The objective of the present study was to identify areas of concern for dairy cow welfare defined at a 'criteria level'. Welfare assessments used to highlight the potential weak points were evaluated at both herd and population level using WQ as a 'gold standard method' and the newly developed DAWIN, respectively.

Materials and methods

This study was conducted as a cross-sectional observational study, with data collected during one visit per herd from January to December 2015 in Denmark. At all herd visits, the Welfare Quality® (WQ) protocol and the newly developed Danish National Animal Welfare Index (DAWIN) protocol were used to evaluate animal welfare.

Observers

Nine observers participated in a one-week WQ training session conducted by a member of the Dairy Cattle working group within the Welfare Quality® Network and two colleagues with experience of using the DAWIN Dairy Cattle protocol. Furthermore, the team of observers re-trained in class and on-farm for three additional days. Agreement of observer scoring for the animal-based measures in the WQ protocol was evaluated according to the official WQ Gold Standard scores, with good agreement ($r > 0.7$) for all WQ measures. This was also the case for the animal-based measures in the DAWIN protocol as these were equal to the WQ measures. Agreement for resource-based measures was considered at on-site training sessions using very detailed descriptions within the DAWIN protocol in order to ensure a common understanding of the relevant categorisation of given resource-based measures.

Study herds

In total, 60 Danish dairy herds participated in the present study. Herds were sampled from the Danish Central Husbandry Register (CHR) based on a stratified sample of all Danish dairy herds. An inclusion criterion of a minimum of 80 cow years (one cow per 365 feeding days) per herd was used in order to be representative of the Danish dairy population which had been increasing in herd size and to ensure a sufficiently large on-farm sample size to calculate robust WQ and DAWIN scores, even at a low measure prevalence. A stratified sampling strategy was used due to the predicted correlation between mortality and animal welfare, using annual cow mortality percentages calculated at herd level as the monthly summed proportion of dead animals (assisted and non-assisted deaths) from the total number of animals present in the herd in a given month. In order to obtain a wider range of welfare levels in the study population, herds were categorised into three strata: low (0 to 3.3%), medium (3.3 to 5.2%) and high mortality (> 5.2%). Based on previous experience of recruiting Danish dairy herds, an initial sample of 40 candidate herds within each stratum was drawn from the CHR register in anticipation that approximately 20 herds from each stratum would ultimately be available for inclusion. As shown in Table 1, this method was not quite fulfilled. Recruitment letters were sent to each of the selected herds with an attached factsheet about the DAWIN project. Farmers were subsequently contacted by telephone to confirm participation and ensure access to the register data needed for the WQ and DAWIN assessment (ie milk recordings and health records). Out of the 120 herds contacted, 30 declined the invitation, while another 30 herds did not respond to either

© 2020 Universities Federation for Animal Welfare

https://doi.org/10.7120/09627286.29.2.197 Published online by Cambridge University Press

Table 1 Descriptive characteristics of the 60 Danish dairy herds included in welfare assessments in 2015.

		Sample population (%)	Study herds (n)	National population (%)[†]
Cow mortality (n = 59)	Low mortality (< 3.3%)	34	20	
	Medium mortality (3.3–5.2%)	36	21	
	High mortality (> 5.2%)	30	18	
Herd size (n = 60)	< 100	7	4	30
	101–200	37	37	42
	201–300	18	11	16
	301–400	7	4	6
	> 400	7	4	6
Farming system (n = 60)	Conventional	85	51	90
	Organic	15	9	10
Breed (n = 60)	Large (Holstein or Danish Red)	78	47	87
	Small (Jersey)	22	11	13
Housing system (n = 60)	Free stall; cubicles	93	56	
	Free stall; deep bedding	5	3	81[‡]
	Tie-stall	2	1	19

[†] Total number of herds; n = 3,165;
[‡] Types of alternative lying areas not specified in national data.

the letter or the telephone invitation. A total of 60 herds agreed to participate in the study, yielding observations from a total of 3,591 cows. The mean herd size was 201 cows (range: 82–724) and annual mean mortality was 4.95%. Both means correspond well to the Danish national means for the year 2015, with a mean herd size of 167 cow years and a mean mortality of 5.0% (SEGES 2015). The distribution of herd characteristics is given in Table 1.

Data collection on farms

Full WQ assessments were carried out using the most recent version of the protocol available (Welfare Quality® Network 2013) including behavioural observations (lying, social, avoidance and qualitative behaviour), clinical assessments, collection of resources and management data. In combination with the DAWIN protocol, a total of 43 measures were assessed for the dairy cows in each herd. A large proportion of the DAWIN measures (n = 13) were recorded in a similar way to the WQ protocol, but not all measures from the WQ protocol were graded in the same manner and some were not included in the DAWIN protocol (Table 2). A full description of the DAWIN protocol can be found in Appendix 1 (see supplementary material to papers published in *Animal Welfare*: https://www.ufaw.org.uk/the-ufaw-journal/supplementary-material).

Herd visits were conducted by two or three observers starting 1 h after the morning feed to ensure a sufficient number of cows present at the feeding table. In order to assess measures from both protocols in the most appropriate sequence, data collection was assigned to observers on-farm, following a strict procedure as shown in Table 2 and starting with behavioural observations ('Avoidance distance' and 'Social behaviour') and resource-based measures observed from outside the animal area. After the social and lying-down behaviour observations had ended, observers then entered the animal area and began observing the remaining behavioural observations (ie pain face and getting-up behaviour), clinical measures and the remaining resource-based measures. The total time for on-farm observations ranged between 6 and 8 h.

Resource-based measures were collected in the barn after all animal-based measures had been assessed. An interview on management practices took place at the farmer's preferred time and her/his written consent for collecting the herds' registry data from the Danish Cattle database was received.

Sampling of animals

Within-herd sample sizes were determined according to the WQ protocol. Cows were selected by a random sample function based on the full list of cows present in the herd prior to the visit and identified in the cow area afterwards by their ear-tag numbers. Sample size calculations were based on an allowable error margin of 10% at a prevalence of 50% with a confidence level of 95% which, for instance, led to a sample size of 49 cows in a herd with a total of 100 cows. The initial sample list was expanded by an additional 10–15% cows depending on the herd size in order to ensure a sufficient herd sample, since cows in sick and calving pens were excluded from the data collection. Sampling for the selected DAWIN measures was not random, as all lying cows were assessed for lying outside of the lying area. Pain face was evaluated for 10% of lactating cows (ie every tenth

https://doi.org/10.7120/09627286.29.2.197 Published online by Cambridge University Press

Table 2 Included measures and the sequence of data collection in the assessment of dairy cow welfare in 60 Danish dairy herds using the Welfare Quality® (WQ) protocol and the Danish Animal Welfare Index (DAWIN) protocol.

Observer position	Measure type	Measure	Protocol
Outside cow area	Behavioural	Avoidance distance at feeding table	WQ + DAWIN
		QBA	WQ
		Segment scan:	
		Cows lying partly/completely outside of lying area	WQ
		Agonistic behaviour	WQ
		Coughing	WQ
		Time needed to lie down	WQ + DAWIN
		Animal colliding with equipment	WQ + DAWIN
Outside cow area	Resources	Cubicles	DAWIN
		Feed bunk space	DAWIN
		Tethering	WQ + DAWIN*
		Access to sick pen	DAWIN
		Animals per sick pen/place	DAWIN
		Access to calving pen	DAWIN
		Animals per calving pen	DAWIN
		Total floor area per animal in sick/calving pen	DAWIN
		Access to pasture/outdoor loafing area	WQ
Inside cow area	Behavioural	Pain face	DAWIN
		Lying outside of the lying area	DAWIN
		Getting-up behaviour	DAWIN
		Time needed to lie down	WQ + DAWIN
	Resources	Water supply	WQ + DAWIN*
		Total floor area	DAWIN
		Bedding material	DAWIN
		Cow brush	DAWIN
	Clinical	Body condition score	WQ + DAWIN
		Cleanliness of cow	WQ + DAWIN
		Integument alterations	WQ + DAWIN
		Overgrown claws	DAWIN
		Diarrhoea	WQ
		Vulvar discharge	WQ
		Hampered respiration	WQ
		Nasal discharge	WQ
		Ocular discharge	WQ
		Lameness	WQ + DAWIN
Interview	Management	Bulk tank somatic cell count	WQ + DAWIN
		Mattress type/age	DAWIN
		Mortality	WQ + DAWIN*
		Stillborn calves	WQ + DAWIN
		Euthanasia procedures	DAWIN
		Dystocia	WQ
		Downer cows	WQ
		Access to pasture/outdoor loafing area	WQ
		Dehorning procedures	WQ

* DAWIN measures scales that differ from WQ scales (see also Appendix 1: https://www.ufaw.org.uk/the-ufaw-journal/supplementary-material).

© *2020 Universities Federation for Animal Welfare*

https://doi.org/10.7120/09627286.29.2.197 Published online by Cambridge University Press

cubicle until the sample size was met), while getting-up behaviour was assessed for a minimum of 15 randomly sampled cows by assessing one cow for every fifth cubicle until the sample size was met.

A human approach test to evaluate the avoidance distance of cows was performed at the feeding table on a random sample of the cows present (eg evaluating every nth cow) until the required sample size was met. For behavioural observations (ie QBA and social behaviour in the WQ protocol), the barn was divided into four to six segments that covered all the functional areas to which cows had access, ie areas related to feeding, drinking, resting, grooming and milking robots (if present) had to be represented. Behavioural segments were assessed in a random order, and the observations were made consecutively through all segments until the total maximum observation time of 2 h was reached. Clinical assessments were performed by two observers at the same visit, with cows selected for inclusion either by allocating each observer to specific sections of the barn or by spray-marking cows after evaluation when assessing cows in the same sections. To balance laterality, determination of the body side to be assessed was pre-defined on the recording sheets.

The DAWIN score

Since DAWIN was developed as a monitoring tool for the whole target population of Danish dairy herds, there was a great emphasis put on the applicability and robustness of the DAWIN score with a number of requirements to be met. The aggregation process needed to be practical and transparent. The DAWIN score model therefore consists of a multi-step aggregation of weighted measures with weights deriving from a Danish expert panel similar to the approach described by Burow *et al* (2013). The final DAWIN score shows a weighted population average of the 29 measure prevalences.

The weighting of the scores was derived from an online survey of 38 Danish dairy experts including 12 veterinary bovine practitioners, 14 official animal welfare auditors, six dairy production advisors and six animal welfare researchers (for more details, see Otten *et al* 2016). However, two of the official animal welfare auditors did not answer any of the questions. Experts were asked to assign weights to the DAWIN measures at their respective graded levels of either normal or moderate or severe deviations. Similar to the WQ approach, scores ranged from 0 (lowest possible welfare) to 100 (highest possible welfare), where scores below 20 indicated very poor welfare, scores between 20–50 indicated impaired welfare and scores above 50 indicated acceptable welfare. Five of the nine graded DAWIN measures (time needed to lie down, collisions, mortality, SCC and avoidance distance) were graded according to thresholds similar to the WQ. The remaining four measures of water supply, animals lying outside/partly outside, getting up and stillborn calves were graded according to expert opinion. The final weighting of the measure scores was calculated as the mean value of all expert scores. As an example, expert scores for the welfare impact on cows with mild integument alterations due to hairless patches (moderate deviation from the optimum) was 49.7, while the score for cows with wounds and swellings (severe deviation from the optimum) was 22. A herd with a total of 80 cows, including 50 cows with no integument alterations (weighted score = 100), 20 cows with moderate integument alterations (weighted score = 49.7), and ten cows with severe integument alterations (weighted score = 22) would give a total measure score for integument alterations of: $(50 \times 100 + 20 \times 49.7 + 10 \times 22)/80 = 77.68$. In order to account for the varying number of cows in the different sections assessed per herd, measure scores were calculated using a weighted average per section. The final population measure score for integument alterations was calculated as the average of all 60 herds.

Data processing and score calculations

Data recording sheets were transcribed from paper to electronic files with a random control for typing errors. The data were subsequently read and subjected to a variety of systematic tests to minimise the possibility of data entry errors using a bespoke package written using the R statistical programming language (R Core Team 2016). The transformation of output measures and aggregation to herd score was also done in R, along with a summary of the measure score distribution and other summary statistics, eg herd prevalence.

WQ criteria, principles and overall classifications were calculated at herd level according to the published procedures (Welfare Quality® 2013) with recent updates. Results were cross-checked with the online Welfare Quality® score calculation provided by INRA (Welfare Quality® 2009). Both WQ and Index scores cannot be calculated in case of missing values, hence, this was handled by imputing population averages for the given measure in a herd with missing values.

Weak point analysis

We used the recommendations given by experts within the WQ protocol (Welfare Quality® 2013) for the weak point analysis. Possible scores of 0–100 could be achieved at criteria and principle level, with < 20 indicating critical or very poor animal welfare, and between 20–40 signifying poor welfare where improvements are necessary and advised. Scores between 40 and 60 indicate a medium level of welfare (not good but not bad), while > 60 is considered to be good to enhanced welfare and > 80 indicates excellent welfare.

The weighted DAWIN scores were classified as problematic (Y) when scores fell below 50. Additionally, the single measures were also classified as problematic according to WQ if measure medians or proportions exceeded the respective measure thresholds given by WQ experts where applicable (given in brackets in Tables 3 and 4; see supplementary material to papers published in *Animal Welfare*: https://www.ufaw.org.uk/the-ufaw-journal/supplementary-material). However, in the case of DAWIN-specific measures (ie water supply, lying outside of lying area, milk SCC, stillborn calves and getting-up behaviour), thresholds from the DAWIN experts were used.

https://doi.org/10.7120/09627286.29.2.197 Published online by Cambridge University Press

At WQ measure level, different units of the parameters required individual thresholds, which were adopted from the same protocol (Welfare Quality® 2013). In the present study, the population median was used to identify problem areas in WQ measures, criteria and principles. This means that even if an area was not identified as a welfare problem at population level, it could still be an individual problem in some herds (and *vice versa*), which could be recognised from the minimum and maximum value for each measure, criterion and principle. WQ population medians were classified as either a 'severe welfare problem' (Y_s) when WQ criterion scores were lower than 20, or a 'mild welfare problem' (Y_m) when WQ criterion scores were greater than 20 but lower than 40. All other WQ population medians were classified as 'no problem' (N).

Results

Descriptive results for measures, criteria and overall welfare scores for the DAWIN protocol are given in Table 3 and for the WQ protocol in Table 4 and both can be found at https://www.ufaw.org.uk/the-ufaw-journal/supplementary-material on the UFAW website.

DAWIN protocol

The overall DAWIN population score for the sample population yielded a median of 77.6 with a very narrow range of 68.0–88.6, indicating similar levels across study herds, and all herds were above the overall critical score of 50 for acceptable welfare. The lowest and most problematic measure scores were found for water supply (40.6), total floor area per section (47.1), cow brushes (47.7) and integument alterations (49.0). However, none of the criteria scores were below 50, and in general there was a relatively small variation found for criteria scores across herds.

The largest variation was found for the criterion 'Absence of prolonged thirst', where herd scores ranged from 30.9–100. The criterion scores for 'Ease of movement' ranged between 45.8 and 100 since they only depended on tethering (score 45.8) or loose-housing (score 100). No variation was found within the resource-based measures of access to calving/sick pen (where all scores were 100) and proper euthanasia. For the animal-based measures, the smallest ranges were found for SCC (85.9–100) and BCS (72.7–100).

WQ protocol

Herds in the present study covered the entire possible range of scores for the criterion 'Comfort around resting', derived from behaviours relating to resting and cleanliness. However, as only one out of the 60 herds was a tie-stall herd, there was no apparent variation in scores for this particular criterion. The full range was also used throughout for all the clinical signs of disease and injuries at measure level, with the highest measure prevalence for lameness (up to 79% moderately lame) and integument alterations (up to 86% with mild alterations). The remaining measures did not exceed a prevalence of > 36%).

For the social behaviours and avoidance distance, the full range of measures was used to a large extent, while access to pasture was restricted, and in most herds even absent. For the first Principal Component (PC1) of the Qualitative Behaviour Assessment (QBA), the potential range of positive and negative emotional states was not covered completely, as scores remained around a middle range, with absence of both extreme positive and negative emotional states.

It was possible for median criterion scores to range from 0–100 (poor to excellent welfare), and this was found to be the case in this study, with both extremes found from measurements for which the study population was fairly uniform and showed almost no variation (ie 'Expression of other behaviour' [0] and 'Ease of movement' [100]). The more differentiated criteria ranged from 26 in 'Comfort around resting' (one farm achieving a score of 89, while others ranged from 0–63), to 82 in 'Expression of social behaviour' (only three farms achieved scores below 47, while others ranged from 54–100). The principle levels 'Good feeding' and 'Good housing' achieved high median scores (60 and 54 of 100, respectively), while at the same time presenting a wide range across farms (scores 9–100 and 2–93, respectively). In contrast, 'Good health' had a low median score (38 of 100) and a limited range (20–76), which was even more pronounced in 'Appropriate behaviour' (median score: 36 of 100; range: 17–64). The overall classification left one farm 'Not classified', 30 farms 'Acceptable' and 29 farms 'Enhanced', yet no farms were graded 'Excellent' in terms of their level of animal welfare.

Weak point analysis DAWIN

The weak point analysis identified eight measures of concern in terms of their impact on dairy cow welfare in the sample population when expressed as crude median measure scores or proportions rather than the weighted DAWIN measure scores in Table 3 (https://www.ufaw.org.uk/the-ufaw-journal/supplementary-material). Measures that exceeded thresholds, thus indicating problematic average welfare levels included: integument alterations, time needed to lie down, animals colliding with equipment, water supply, cow brushes, total floor area and SCC.

The most prominent welfare issue identified by the DAWIN protocol was 'Absence of prolonged thirst', which affected the largest proportion of herds and animals within the sample population since only 15.8% of the herds had a sufficient water supply, while 40.3% of the herds had an insufficient water supply. In this case, thresholds were based on the presence of no more than six cows per drinker or at least a 10-cm trough per cow. This requirement was not met in almost half of the herds (47.9%). A total of 140 barn sections in the sampled herds were assessed. Out of these, 95 sections in 55 herds had an insufficient water supply according to the protocols due to too few drinkers in 10.5%, too short trough length in 74.3%, a combination of both in 11.6% and no available drinkers/troughs in 3.6% of the sections.

Other areas of concern included the insufficient number of rotating cow brushes since only 11.9% of the herds were equipped with at least one rotating cow brush per 50 animals. Only 24.6% of the herds met the minimum

© 2020 Universities Federation for Animal Welfare

https://doi.org/10.7120/09627286.29.2.197 Published online by Cambridge University Press

breed-specific requirements for a total floor area per animal of 8 m^2 for large breeds and 6.6 m^2 for small breeds. Finally, 11.25% of the study herds exceeded the threshold for the proportion of cows with a chronically high SCC (> 400,000 cells ml^{-1}). Almost all of these herds were classified as 'problematic', with only two herds within the range of 'moderate problematic' levels (between 2.25–4.5% of cows with elevated SCC) and none within the 'not problematic' range (< 2.25% of cows with elevated SCC).

Weak point analysis Welfare Quality®

The in-depth analysis of WQ outcomes identified weak points of animal welfare at herd level (Y_m and Y_s in Table 4) for two principles, three criteria and 17 single measures. Although the principle 'Good feeding' indicated a good overall level of welfare, the median number of water points within the criterion 'Absence of prolonged thirst' highlighted a mild welfare problem according to the WQ protocol. A mild problem can be interpreted as the available trough length being under 6 cm per cow and/or one bowl serving as a water point for more than ten animals on a typical farm.

Overall, the principle of 'Good housing' indicated a good level of welfare, yet cows on a typical farm were subjected to welfare issues relating to the criterion 'Comfort around resting'. The deficits were moderate in terms of the measure mean time to lie down, referring to a prolonged duration between 5.2 and 6.3 s. Furthermore, cows lay outside of the cubicles in 3.5% of observations over the 2-h observation period. Severe problems were identified in colliding with equipment when lying down (> 30% of the cows). Furthermore, severe welfare impairments were identified in relation to the cleanliness of rear body parts, indicating a prevalence of > 50% of cows with dirty lower legs and > 19% with dirty udders, soiled flanks and upper legs.

The principle 'Good health' was also identified as an area of poor welfare. A closer look at the associated criteria revealed that within 'Absence of injuries', population medians of the associated measures integument alterations and lameness were above the thresholds for mild welfare problems. The criterion scores for 'Absence of disease' indicated a neutral level of welfare (WQ scores 40–60). However, three of the ten included measures indicated problem areas: severe impairments for elevated SCC and dystocia and mild impairments for mortality. 'Absence of pain induced by management procedures', the third criterion within this principle, also indicated neutral welfare, yet disbudding with the thermal method and a widely neglected use of analgesics pointed towards a welfare problem in this criterion and principle ('Good health').

A severe deficit was identified in the criterion 'Expression of other behaviour', measured as access to pasture over a year, which also led to a mild welfare problem at principle level for 'Appropriate behaviour'.

Discussion

The objective of the present study was to identify weak points in dairy cow welfare at both herd and population level in Denmark. As the construct of the newly developed DAWIN protocol for dairy cattle was similar to the WQ cattle protocol, there was an overlap in areas of concern found when using both approaches, such as water supply, resting comfort and chronically elevated SCCs.

Animal welfare in the sampled herds and in the national population

The findings of the present study showed that average scores for the included measures indicated acceptable levels of animal welfare for Danish dairy cows. Only one farm received a 'Not classified' grade in the WQ, and no farms scored below the acceptable threshold of 50 for the DAWIN. The DAWIN score at herd level should be treated with caution since there is perfect compensation, ie a low welfare score in one measure can be compensated for by a high welfare score in another (this is not the case for WQ, for a discussion of different aggregation procedures, see Sandøe *et al* 2019). Direct extrapolation to the entire Danish population should be carried out with caution. Herds were sampled based on the inclusion criterion of having more than 80 cows, and further stratified based on cow mortality risk, which is not completely representative of the Danish dairy population. Although the study herds are to some extent representative as they cover the herd characteristics needed to meet future challenges within dairy production, some important types of herds are either sparsely represented or not represented at all, eg smaller herds with fewer than 100 cows and/or the tie-stall herds. However, it can be argued that these two types are related as many of the small-scale herds are housed in tie-stalls, accounting for only 20% of dairy production sites in Denmark. Furthermore, tie-stall barns will be completely banned in Denmark from 2027, and since the increase in marginal revenue is proportional to herd size, the future of these small-scale production units might be threatened.

Herds in the present study were recruited in a similar way to previous studies evaluating welfare in Danish dairy herds at herd level. The inclusion criteria herd size and loose-housing with at least 100 cows (Burow *et al* 2013; Otten *et al* 2016) or 50 cows (Andreasen *et al* 2014) have been used, while the latter study also included breed as an additional inclusion criterion. In all of these studies, lists with potential study herds were requested from and supplied by the Danish Cattle Federation. In the present study, herds were identified in the Central Husbandry Register (CHR), a national, publicly available database containing information about herd size, breed, and production system (organic or conventional). Information on housing options was not supplied, leading to the inclusion of tie-stall herds, which was in contrast to the previous Danish studies. Despite the different sampling strategies, the results and conclusions relating to the welfare of Danish dairy cows were similar in this and previous studies.

https://doi.org/10.7120/09627286.29.2.197 Published online by Cambridge University Press

Table 5 Comparison of European studies evaluating animal welfare at herd level using Welfare Quality®.

Study	Country	Number of farms	Mean herd size	Herd WQ classification (%)
Ostojic-Andric *et al* (2011)	Serbia	LH = 3 TS = 3	LH = 56 TS = 77	LH: NC=0/AC=0/EN=100/EX=0 TS: NC=0/AC=33.3/EN=66.7/EX=0
Andreasen *et al* (2013)	Denmark	LH = 43/TS = 0	184	NC=2.3/AC=46.5/EN=51.2/EX=0
de Vries *et al* (2013)	The Netherlands	179	67	NC=9/AC=47.4/EN=43.6/EX=0
Popescu *et al* (2014)	Romania	LH = 30 TS = 30	LH = 84 TS = 70	LH: NC=20/AC=63.3/EN=16.7/EX=0 TS: NC=0/AC=30/EN=70/EX=0
des Roches *et al* (2014)	France	131	51	NC=4.6/AC=57.3/EN=36.6/EX=0
Tremetsberger *et al* (2015)	Austria	LH = 34/TS = 0	35	NC=2.9/AC=41.2/EN=55.9/EX=0
Krug *et al* (2015)	Portugal	24	80	NC=20/AC=75/EN=5/EX=0
DAWIN project	Denmark	LH = 59/TS = 1	201	NC=1.7/AC=50/EN=48.3/EX=0
Gieseke *et al* (2018)	Germany	80	383	NC=4/AC=66/EN=30/EX=0

LH = Loose housing; TS = Tie-stall;
Welfare Quality® classification: NC = Not classified; AC = Acceptable; EN = Enhanced; EX = Excellent.

Problem areas identified by DAWIN and previous Danish studies

The transparent index approach using weighted sums has previously been used in Danish settings for pigs and cattle. Burow *et al* (2013) investigated differences in integrated welfare indices in dairy herds at the end of the winter barn period compared to the end of the grazing period. Pig indices were created by Knage-Rasmussen *et al* (2015) and dairy cattle indices by Otten *et al* (2016) in order to evaluate the potential of remote welfare assessment by means of routinely collected farm data. Both studies compared indices based on routinely collected register data to indices based on on-farm assessments. The previous two cattle studies found a spread in overall index scores across herds covering 54 and 49% of the possible range of scores, whereas our study showed a markedly lower spread with only 18.8% (Index scores 68.0–86.8) of the possible score range being used across herds. An explanation for this difference might be that measures in the DAWIN are a combination of resource- and animal-based measures, while the other two studies relied more heavily on animal-based measures. Likewise, the aggregation procedures were not identical, as the DAWIN adjusted for herd size and used different expert weightings. Nonetheless, mean index scores for all three index studies, including the DAWIN, indicated neutral welfare, as Burow *et al* (2013) reached a mean of 2,926/5,400 index units, corresponding to 54% of the maximum score, and Otten *et al* (2017) reached a mean of 33.5/52 index units, corresponding to 64% of the maximum score. The higher mean score in the DAWIN (80.5%) might also be attributed to the implementation of The Act on Keeping of Dairy Cattle and Their Offspring (2010). The two previous studies were conducted during 2010, and a number of legislative incentives concerning resource-based features in cow barns were not implemented until the end of a transition period, meaning that the first effective implementations were enforced from 2016 for barns built prior to the act being passed in 2010. The positive welfare effects of these implementations might also be mirrored by the higher mean DAWIN score, as the resource-based measures showed the smallest variation in scores across herds, which might be due to farmer awareness of compliance and, likewise, on the production-related outcomes represented by SCC and BCS.

Comparison of Danish WQ farm results and other European WQ farm studies

At present, WQ data have been gathered in at least ten European countries from 491 herds (de Graaf *et al* 2017) in addition to our 60 herds. Comparisons among the European results are not straightforward, as they may be influenced by the differences in sampled units and the variation in agricultural structures across countries. Differences in sampled herds, herd sizes and WQ classifications are depicted in Table 5, which shows, eg a larger proportion of small-scale dairy units in most of the previous studies, in which mean herd sizes were predominantly under 100 cows, compared to the larger dairy units in both the present and previous Danish studies. The similar study design of the two Danish studies also resulted in a greater agreement in classification patterns. The WQ results at the highest level of aggregation within this study (the farm score) can be compared to the results found by Andreasen *et al* (2013; see Table 5). A study assessing dairy herds in the northern part of Germany (Gieseke *et al* 2018) showed similar classification patterns and is likely to be most comparable to Danish production settings. However, only herds practicing summer grazing were included, in contrast to the present study.

The health and welfare levels assessed in the reported studies of Danish dairy herds correspond well with comprehensive study publications available from other European countries using the same WQ assessment protocol or slightly modified versions of it. At measure level, in particular, many studies

© 2020 Universities Federation for Animal Welfare

https://doi.org/10.7120/09627286.29.2.197 Published online by Cambridge University Press

enable comparisons among parameters related to lying comfort, clinical remarks and behaviour. Amongst these measures, body condition in Danish dairy cows has been scored according to the WQ score in several previous studies. Gratzer (2011) assessed Danish herds within a European study reporting a median of zero percent very lean cows, a finding also reported by other authors describing rather low percentages of very lean cows, with medians of 0.0 for winter or 3.2% for summer (Burow *et al* 2013) and 2% (Rousing *et al* 2013), respectively. In contrast, our study found not only a higher median of 5.22, but also a higher maximum of 32.14. Likewise, a recent study by Gieseke *et al* (2018) also showed higher mean percentages of lean cows between 9–17.7% depending on herd size and season. On an international level, Ostojic-Andric *et al* (2011) reported an average of 5% for loose-housed cows in Serbia, Blanco *et al* reported a median of 4.57% in organic farms in Spain (I Blanco, personal communication 2014), and Tremetsberger *et al* (2015) reported means between 16–24% in different farm groups in an Austrian intervention study. This shows that although the proportion of lean cows in Danish dairy herds might have risen in recent years, it is generally still low compared to international levels. This might be due to the relatively high amount of TMR feeding in Danish dairy farms, and the presence of cow transponders allowing additional concentrate to be delivered based on milk yield. Nonetheless, as the WQ threshold suggested, lean cows might still present a welfare problem at cow and farm level, as indicated by the range of affected cows.

In addition, cows in the present study needed a longer mean time to lie down than recommended (5.8 s, maximum: 11.31) compared to results found in previous European studies. Gratzer *et al* (2011) reported 4.0 s in Denmark, 4.4 s in the UK, 5.0 s in Switzerland, 5.3 s in The Netherlands and Austria, and 5.5 s in Germany, compared to the 5.9 s found more recently by Gieseke *et al* (2018). In a study by Kirchner *et al* (2014), dairy cows had a median lying down time of 5.08 s in Northern Ireland, 4.03 s in Spain and 4.29 s in Romania, indicating shorter durations than the present results. One study (Tremetsberger *et al* 2015) reported 6.0 s in a control group of Austrian dairy cows in the initial year. All these studies used the same trainers, training materials and observer testing, however, the cows themselves might be a source of the variation encountered. Firstly, cows display different laying down behaviour in the different housing systems present across the studies. Secondly, it is important to consider that the minimum number of cows scored is six, and although this was validated earlier (WQ report 2009b), the influence of prolonged laying down for individual animals cannot be completely excluded. Nonetheless, the expert rating within WQ and the comparison on international level both highlight a deficit.

Other areas of concern indicated at measure level can be found and compared on a national and international level based on previously mentioned studies. Examples include lying outside of the cubicles (Tremetsberger *et al* 2015); colliding with equipment when lying down (Kirchner *et al* 2014); cleanliness of rear body parts; integument alterations (Burow *et al* 2013; Rousing *et al* 2013) and lameness (Gratzer *et al* 2011; Burow *et al* 2013; Rousing *et al* 2013; Kirchner *et al* 2014; Tremetsberger *et al* 2015). However, a comprehensive comparison of these measure results should be performed. Nevertheless, the WQ results of the present Danish study resemble previously reported studies using the WQ protocol.

Areas of concern identified by WQ and DAWIN

The welfare assessments in the present study were aimed at different targets (ie herd level and national level), and the aggregation approaches therefore differed not only in the weightings assigned by experts, but also in the different aggregation steps. Nonetheless, final results highlight an overlap in the major problem areas identified in Danish dairy herds, as all four overarching welfare principles show weak points.

Water supply was an area of concern in both assessment approaches, which is also in line with the findings of the official welfare audits performed by the DVFA. Since 2008, welfare inspections have been performed within a risk-based scheme, where 5% of all livestock herds with more than ten animals are subjected to animal welfare compliance controls, meaning that 500–700 cattle (ie beef and dairy) herds are visited per year. Non-compliance issues, in particular regarding insufficient feed and water supply and/or improper handling of sick and injured animals, are usually found in 15–25% of the herds, with more severe issues leading to police reports in 1.5–2% of the inspected herds. The thresholds for determining a sufficient water supply in both the WQ and DAWIN protocol are somewhat similar to the Danish legislation. As such, a high proportion of farms in the present study not only fail to provide a water supply in terms of welfare assessment schemes, but also in terms of compliance with the legislation. Current requirements are for a maximum of six cows per drinker according to The Danish Act on Keeping of Dairy Cattle and their Offspring (LBK no 58 11/01/2017), as implemented in July 2016. However, the requirement for a minimum of a 10-cm trough per cow remains in a transition period until July 2024. Insufficient water supply is not only a Danish welfare issue as studies in Germany (Gieseke *et al* 2018; Wagner *et al* 2018) and Belgium (de Graaf *et al* 2017) indicated lower WQ criterion scores for 'Absence of prolonged thirst' for herds during summer grazing. These findings are also in accordance with Danish results found by Burow *et al* (2013). However, none of the sample herds in the present study were assessed during the grazing period, which actually emphasises the concern for Danish herds as the water supply is impaired at all times. Similar results were found in German dairy herds, where WQ criterion scores for water supply were significantly lower in smaller herds, lowest in herds with 100–299 cows compared to herds below 100 cows (Gieseke *et al* 2018).

Measures relating to resting comfort (time needed to lie down, collisions with equipment and cleanliness) were also problematic based on both assessment protocols due to the similar assessment and classification in both the WQ and the DAWIN. While the WQ assessment highlighted

https://doi.org/10.7120/09627286.29.2.197 Published online by Cambridge University Press

lameness and integument alterations at problematic levels, the DAWIN did not provide any thresholds to define problematic proportions for these measures.

The problematic welfare impact of the WQ principle 'Appropriate behaviour' was due to the generally poor access to grazing amongst study herds driving the criterion score in this particular protocol. Access to pasture was not included in the DAWIN protocol for numerous reasons. In Danish production settings, available farmland close to cow barns is scarce and cows must walk long distances on stony and/or muddy tracks between milking and grazing, increasing the risk of lameness (Somers *et al* 2003; Burow *et al* 2014), as well as increasing man hours and labour. Although access to pasture has proven to be beneficial for other welfare aspects such as hock lesions (Burow 2013), claw conformation and lameness (Corrazin *et al* 2010), it was excluded from The Danish Act on Keeping of Dairy Cattle and their Offspring (LBK no 58 11/01/2017). To compensate, cows were ensured more space indoors, with a minimum space requirement of 8 m^2 or 6.6 m^2 per cow. However, DAWIN also indicated space requirements as an area of concern, as almost a quarter of the study herds did not meet the minimum space requirements. The higher stocking density may have different causes, eg it might be a consequence of the withdrawal of milk quota in 2015, which enabled farmers to increase income by enhancing the milk yield. The DAWIN used the measure cow brushes to cover 'Appropriate behaviours', which also revealed severe deficits within the herds. Hence, the overall welfare issues identified by both protocols were similar at least at criteria level, even if they were caused by different underlying measures.

Animal welfare implications and conclusion

This study pinpointed major welfare concerns in Danish dairy herds based on two different approaches: at herd level using the WQ protocol, and at national level using the DAWIN protocol. Despite the differences in both the quantity of measures and their aggregation into final scores, both approaches showed an overall agreement in terms of the areas of concern. The most crucial concern was insufficient water supply, which was identified by both approaches for the criteria 'Absence of prolonged thirst'. Both approaches also identified similar welfare deficits for the criteria 'Comfort around resting via thresholds' for the mean time needed to lie down, collisions with equipment, lying outside of the designated lying area and cleanliness of rear body parts. At herd level, the WQ also identified the directly associated measures lameness and integument alterations within the criterion 'Absence of injuries' as consequences of impaired resting comfort. These animal-based levels were not found to be problematic at national level, however, the resource-based measure total floor area was a welfare issue. Both approaches agreed on the measure of SCC being a welfare concern, while dystocia was only identified as a problem at herd level. Finally, the 'Expression of other behaviours' was found to be problematic in both approaches, but these deficits were determined by different measures as the WQ defined them by a lack of access to pasture and the DAWIN by an insufficient number of cow brushes present in the barns. In conclusion, the DAWIN protocol uncovered welfare issues in a valid manner and in accordance with the WQ results found not only on a national but also on a European level.

Acknowledgements

This study was funded by the Ministry of Environment and Food of Denmark and the authors wish to acknowledge Louise Holm Parby, Else Enemark and Dorte Schrøder from the Danish Veterinary and Food Administration for their contribution in the project and Pia Haun Poulsen and Phillipe Deborde from Aarhus University for their help with farmer recruitment and the early data collection phase. We also wish to express our gratitude to the participating farmers for opening their doors and making studies such as these possible.

References

Andreasen S, Sandøe P and Forkman B 2014 Can animal-based welfare assessment be simplified? A comparison of the Welfare Quality® protocol for dairy cattle and the simpler and less time-consuming protocol developed by the Danish Cattle Federation. *Animal Welfare 23*: 81-94. http://dx.doi.org/10.7120/09627286.23.1.081

Andreasen S, Wemelsfelder F, Sandøe P and Forkman B 2013 The correlation of Qualitative Behavior Assessments with Welfare Quality® protocol outcomes in on-farm welfare assessment of dairy cattle. *Animal Welfare 143(1)*: 9-17. http://dx.doi.org/10.1016/j.applanim.2012.11.013

Appleby MC and Sandøe PT 2002 Philosophical debate on the nature of well-being: Implications for animal welfare. *Animal Welfare 11(3)*: 283-294

Bartussek H 1999 A review of the animal needs index (ANI) for the assessment of animals' well-being in the housing systems for Austrian proprietary products and legislation. *Livestock Production Science 61*: 179-192. https://doi.org/10.1016/S0301-6226(99)00067-6

Blokhuis HJ 2008 International cooperation in animal welfare: The Welfare Quality® project. *Acta Agriculturae Scandinavica Section A, Animal Science 50(1)*: S10. https://doi.org/10.1186/1751-0147-50-S1-S10

Burow E, Rousing T, Thomsen PT, Otten ND and Sørensen JT 2013 Effect of grazing on the cow welfare of dairy herds evaluated by a multidimensional welfare index. *Animal 7(5)*: 834-842. https://doi.org/10.1017/S1751731112002297

Burow E, Rousing T, Thomsen PT and Sørensen JT 2014 Track way distance and cover as risk factors for lameness in Danish dairy cows. *Preventive Veterinary Medicine 113(4)*: 625-628. https://doi.org/10.1016/j.prevetmed.2013.11.018

Corazzin M, Piasentier E, Dovier S and Bovolenta S 2010 Effect of summer grazing on welfare of dairy cows reared in mountain tie-stall barns. *Italian Journal of Animal Science 9(3)*: 304-312. https://doi.org/10.4081/ijas.2010.e59

Danish Act on the Keeping of Dairy Cattle and their Offspring 2010 *Lov om hold af malkekvæg og afkom af malkekvæg Lovbekendtgørelse nr 58*. Danish Ministry of Justice: Copenhagen, Denmark

Danish Act on the Keeping of Dairy Cattle and their Offspring 2017 *Lov om hold af malkekvæg og afkom af malkekvæg Lovbekendtgørelse nr 58*. Danish Ministry of Justice: Copenhagen, Denmark

© *2020 Universities Federation for Animal Welfare*

https://doi.org/10.7120/09627286.29.2.197 Published online by Cambridge University Press

Danish Cattle Federation 2005 *The Danish Cattle Federation, business guidelines concerning animal welfare, Kvæginfo 1547 – Dansk Kvægs branchepolitik for dyrevelfærd*. https://www.landbrugsinfo.dk/Kvaeg/Sundhed-og-dyrevelfaerd/Dyrevelfaerd/Sider/Dansk Kvaegs_Branchepolitik_for_Dyrevelf.aspx

Danish Veterinary and Food Administration 2018 *Animal welfare control.* https://www.foedevarestyrelsen.dk/Leksikon/Sider/Velf%C3%A6rdskontrol.aspx

De Graaf S, Ampe B and Tuyttens FAM 2017 Assessing dairy cow welfare at the beginning and end of the indoor period using the Welfare Quality® protocol. *Animal Welfare 26(2)*: 213-221. https://doi.org/10.7120/09627286.26.2.213

des Roches AD, Veissier I, Coignard M, Bareille N, Guatteo R, Capdeville J, Gilot-Fromont E and Mounier L 2014 The major welfare problems of dairy cows in French commercial farms: an epidemiological approach. *Animal Welfare 23(4)*: 467-478. https://doi.org/10.7120/09627286.23.4.467

de Vries M, Bokkers EAM, van Shaik G, Botreau R, Engel B, Dijkstra T and de Boer IJM 2013 Evaluating results of the Welfare Quality® multi-criteria evaluation model for classification of dairy cattle welfare at the herd level. *Journal of Dairy Science 96(10)*: 6264-6273. https://doi.org/10.3168/jds.2012-6129

Fraser D, Weary DM, Pajor EA and Milligan BN 1997 A scientific conception of animal welfare that reflects ethical concerns. *Animal Welfare 6*: 187-205

Gieseke D, Lambertz C and Gauly M 2018 Relationship between herd size and measures of animal welfare on dairy cattle farms with free-stall housing in Germany. *Journal of Dairy Science 101*: 1-15. https://doi.org/10.3168/jds.2017-14232

Gratzer E 2011 *Animal health and welfare planning in Austrian organic dairy farming*. University of Natural Resources and Life Sciences (BOKU), Vienna, Austria

Johnsen PF and Sandøe P 2001 Assessment of farm animal welfare at herd level: Many goals, many methods. *Acta Agriculturae Scandinavica Section A, Animal Science 30*: 26-33. https://doi.org/10.1080/090647001316923027

Keeling L 2009 *An overview of the development of the Welfare Quality® Project Assessment Systems.* Welfare Quality® Reports no 11, Cardiff University, UK

Kirchner MK, Ferris C, Abecia L, Yanez-Ruiz D, Pop S, Voicu I, Dragomir C and Winckler C 2014 Welfare state of dairy cows in three European low-input and organic systems. *Organic Agriculture 4(4)*: 309-311. https://doi.org/10.1007/s13165-014-0074-2

Knage-Rasmussen KM, Rousing T, Sørensen JT and Houe H 2015 Assessing animal welfare in sow herds using data on meat inspection, medication and mortality. *Animal 9(3)*: 509-515. http://dx.doi.org/10.1017/S1751731114002705

Krug C, Haskell MJ, Nunes T and Stilwell G 2015 Creating a model to detect dairy cattle farms with poor welfare using a national database. *Preventive Veterinary Medicine* 3: 280-286. https://doi.org/10.1016/j.prevetmed.2015.10.014

Michelsen AM, Hakansson F, Pedersen-Lund V, Kirchner MK, Otten ND, Denwood M, Rousing T, Houe H and Forkman B Identifying areas of animal welfare concern in Danish pig herds using the Danish Animal Welfare Index (DAWIN). *Animal Welfare,* submitted

Ostojic-Andric D, Hristov S, Novakovic Z, Pantelic V, Petrovic M, Zlatanovic Z and Niksic D 2011 Dairy cows welfare quality in loose vs tie housing system. *Biotechnology in Animal Husbandry 27*: 975-984. https://doi.org/10.2298/BAH1103975O

Otten ND, Rousing T and Forkman B 2017 Influence of professional affiliation on expert's view on welfare measures. *Animals 7(85)*. https://doi.org/10.3390/ani7110085

Otten ND, Rousing T, Thomsen PT, Houe H and Sørensen JT 2016 Comparison of animal welfare indices in dairy herds based on different sources of data. *Animal Welfare 25*: 207-215. https://doi.org/10.7120/09627286.25.2.207

Popescu S, Borda C, Diugan EA, Niculae M, Stefan R and Sandru C 2014 The effect of the housing system on the Welfare Quality® of dairy cows. *Italian Journal of Animal Science 13(1*https://doi.org/10.4081/ijas.2014.2940

R Core Team 2016 *R: A language and environment for statistical computing*. R Foundation for Statistical Computing: Vienna, Austria. http://www.R-project.org/

Rousing T, Thomson PT, Sørensen JT, Otten N and Houe H 2013 Nødvendigt med flere mål for at vurdere dyrevelfærden i en malkekvægbesætning. *Dansk Veterinaer Tidskrift 2*: 14-16. [Title translation: More measures are needed to evaluate animal welfare in dairy herds]

Sandøe P, Corr SA, Lund TB and Forkman B 2019 Aggregating animal welfare indicators: can it be done in a transparent and ethically robust way? *Animal Welfare 28*: 67-76. https://doi.org/10.7120/09627286.28.1.067

SEGES 2015 *Opgørelse vedr kodødelighed mv*. https://www.landbrugsinfo.dk/Kvaeg/Tal-om-kvaeg/Sider/Kodod201508.aspx. [Title translation: Cow mortality figures]

Somers JGCJ, Frankena K, Nordhuizen-Stassen EN and Metz JHM 2003 Prevalence of claw disorders in Dutch dairy cows exposed to several floor systems. *Journal of Dairy Science 86(6)*: 2082-2093. https://doi.org/10.3168/jds.S0022-0302(03)73797-7

Tremetsberger L, Leeb C and 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*: 6801-6811. https://doi.org/10.3168/jds.2014-9084

Wagner K, Brinkmann J, March S, Hinterstoißer P, Warnecke S, Schüler M and Paulsen HM 2018 Impact of daily grazing time on dairy cow welfare. Results of the Welfare Quality® Protocol. *Animals 8*: 1. https://doi.org/10.3390/ani8010001

Webster AJF, Main DCJ and Whay HR 2004 Welfare assessment: indices from clinical observation. *Animal Welfare 13*: S93-98

Welfare Quality® 2009a Simulate results. http://www1.clermont.inra.fr/wq/index.php?id=simul&new=1

Welfare Quality® 2009b *Welfare Quality® assessment protocol for cattle*. Welfare Quality® Consortium: Lelystad, The Netherlands

Welfare Quality® 2013 *Welfare Quality® assessment protocol for cattle.* http://www.welfarequality.net/media/1017/cattle_protocol_without_veal_calves.pdf

https://doi.org/10.7120/09627286.29.2.197 Published online by Cambridge University Press