The Old School, Brewhouse Hill, Wheathampstead,
Hertfordshire AL4 8AN, UK
www.ufaw.org.uk

Animal Welfare 2017, 26: 213-221
ISSN 0962-7286
doi: 10.7120/09627286.26.2.213

Assessing dairy cow welfare at the beginning and end of the indoor period using the Welfare Quality® protocol

S de Graaf†, B Ampe† and FAM Tuyttens*†‡

† Animal Sciences Unit, Institute for Agricultural and Fisheries Research (ILVO), Scheldeweg 68, 9090 Melle, Belgium
‡ Faculty of Veterinary Medicine, Ghent University, Salisburylaan 133, 9820 Merelbeke, Belgium
* Contact for correspondence and requests for reprints: frank.tuyttens@ilvo.vlaanderen.be

Abstract

The Welfare Quality® (WQ) project developed protocols as international standards for farm animal welfare assessment. For dairy cattle, the WQ protocol may be performed any time during the indoor period when the herd has no access to pasture. However, timing of welfare assessments during the indoor period might influence the outcome, as pasture access influences many welfare aspects and such effects are likely to carry over to the beginning of the indoor period. In order to test this hypothesis, we assessed ten herds at both the beginning and end of the same indoor period. Assessment at the end of the indoor period resulted in a higher prevalence of the following welfare measures: severely lame cows, mild integument alterations, coughs per cow per hour, diarrhoea, and vulvar discharge. In addition, at the end of the indoor period, the mortality rate was higher and the Qualitative Behaviour Assessment (QBA) score was worse, but there were fewer cows with dirty udders and mean time to lie down was lower. This led to worse scores on the criterion and principle levels of WQ integration, but resulted in a lower welfare categorisation for one of the herds only. The better scores for several aspects of health and for the QBA at the beginning of the indoor period not only confirm the carry-over of positive effects of access to pasture, but also imply a need for careful consideration of the timing and frequency of WQ assessments of herds that are housed outdoors for part of the year.

Keywords: *animal welfare, dairy cattle, pasturing, welfare assessment, Welfare Quality®*

Introduction

Previous studies have explored possible links between access to pasture and various aspects of dairy cattle welfare. Predominantly positive relationships were found, with access to pasture being related to a lower prevalence of lameness (Onyiro & Brotherstone 2008), a lower incidence of clinical mastitis (Washburn *et al* 2002), and a lower mortality rate (Burow *et al* 2011). When cows are kept indoors for several months during winter (as is common practice in Belgium and other countries with a similar climate), the positive effects of previous access to pasture are likely to carry over to (at least) the beginning of the indoor period, because welfare problems generally take some time to develop. There is little knowledge, however, about how the timing of welfare assessment during the indoor period affects the assessment's outcome. Corazzin *et al* (2010) found that in herds with access to pasture during summer-time, the prevalence of lameness, injuries, coughs and vulvar discharge increased from three weeks after transferral to indoor housing onwards, as did the time it took the animals to stand up after lying down. Burow *et al* (2013) used an aggregated welfare index to assess the welfare of loose-housed Danish herds (all with access to pasture in summer) during the indoor and outdoor periods. They concluded that welfare was better during summer than during winter. This suggests that timing of assessment is critical to welfare assessment of herds that are allowed outdoors for part of the year.

The Welfare Quality® (WQ) protocol for dairy cattle (Welfare Quality® 2009) describes an elaborate procedure to assess dairy cattle welfare. This protocol is one of the few that have been developed to assess overall welfare in addition to separate welfare indicators. It describes 33 measures of dairy cattle welfare as well as a method to integrate these into an overall welfare categorisation: 'excellent', 'enhanced', 'acceptable' or 'not classified'. This protocol is carried out during the indoor period because some of the evaluations are impossible to perform while the animals are at pasture, but the protocol provides no guidelines or restrictions regarding the timing of assessments during the indoor period (ie, beginning, middle or end). However, no studies have been performed to determine whether the timing of such assessments can result in variation of the welfare assessment. Especially for cattle kept indoors during winter only, it is possible that a welfare assessment performed at the beginning of the indoor period may not be comparable with one performed at the end.

To investigate whether the time lapse between the last day of pasture access and the day of assessment actually affects

Table 1 Per farm dates (dd-mm-yyyy) of the first and second farm visits, assessor numbers per farm visit, the period between visits in days, herd size, number of days access to pasture a year and average hours access to pasture a day.

Farm number	First visit		Second visit		Days between visits	Herd size	Days pasture per year	Hours at pasture per day
	Date	Assessor	Date	Assessor				
1	08-11-2010	1 & 2	17-02-2011	3 & 4	101	46	214	12
2	09-11-2010	2 & 3	18-02-2011	4 & 3	101	47	214	11
3	10-11-2010	2 & 1	28-02-2011	1 & 2	119	55	214	5
4	11-11-2010	4 & 3	09-03-2011	3 & 4	109	47	214	7
5	12-11-2010	3 & 4	03-03-2011	4 & 3	111	80	214	8
6	12-11-2010	2 & 4	10-03-2011	2 & 1	118	65	214	8
7	15-11-2010	4 & 3	08-03-2011	2 & 3	113	52	199	5
8	16-11-2010	2 & 3	24-02-2011	3 & 4	100	53	214	13
9	22-11-2010	3 & 4	01-03-2011	4 & 3	99	35	184	6
10	07-12-2010	4 & 3	14-02-2011	2 & 4	69	76	214	7

Table 2 All principles, the corresponding criteria and measures used in the Welfare Quality® assessment protocol for dairy cattle welfare.

Principles	Criteria	Measures
Good feeding	Absence of prolonged hunger	Body Condition Score (% very lean animals)
	Absence of prolonged thirst	Availability* and cleanliness water
Good housing	Comfort around resting	Lying down duration; % collisions; on edge of lying area; cleanliness
	Thermal comfort	No measure for dairy cattle
	Ease of movement	Tethering*
Good health	Absence of injuries	Lameness; integument alterations
	Absence of disease	Respiration/digestive diseases; Somatic Cell Count; mortality; dystocia; downer
	Absence of pain induced by management procedures	Mutilations: dehorning and tail-docking; use of anaesthetics/analgesics
Appropriate behaviour	Expression of social behaviour	Incidence of agonistic interactions
	Expression of other behaviours	Access to pasture*
	Good human-animal relationship	Avoidance distance at feeding place
	Positive emotional state	Qualitative Behavioural Assessment

* Measures cannot differ over time when no significant changes have been made to the farm.

the outcome of single welfare measures and integrated scores, we carried out ten WQ assessments at the beginning of the indoor period and assessed the same ten herds again at the end. We expected the positive effect of previous access to pasture on welfare to decrease after the beginning of the indoor period, thus resulting in better welfare scores at the beginning of the indoor period than at the end.

Materials and methods

Ten herds were assessed both at the beginning (November 2010) and at the end (February and March 2011) of the same indoor period (Table 1). The farms were a convenience sample of the entire Flemish (the northern region of Belgium) population, and were selected based on housing system (loose-cubicle housing), building year (after 1990), herd size (35–120 cows), access to pasture during the warmer period of the year, and farmer consent. The mean (± SD) herd size (55.6 [± 14.1] cows; range: 35–80) was in line with the average Flemish dairy farm (47.5; Federale Overheids Dienst Economie 2016). All herds had access to pasture for ≥ 5 h per day during the outdoor period (mean: 8.2 h; range: 5–13 h). On average, herds had access to pasture for 212.5 days per year (range: 184–244). During the farm visits, animal welfare was assessed using the Welfare Quality® protocol for dairy cattle (WQ 2009). Assessments commenced shortly after morning milking, and were always performed by two out of four trained assessors on the same day. All measures were performed simultaneously, but not necessarily the same sample of cows was observed by both assessors. Exceptions were the clinical scoring measures, where observers did score the same animal independently and simultaneously. The average score of the two assessors was used for analysis, except for measures for which we expected no observer

https://doi.org/10.7120/09627286.26.2.213 Published online by Cambridge University Press

effects potentially biasing the results (ie those derived from the management questionnaire and the resource checklist). The WQ protocol training of the observers lasted three days: two days of theoretical and practical sessions, and one day on-farm. Training was given by an experienced assessor, using the WQ training material.

Welfare Quality® protocol

Only a brief description of the WQ protocol is given here; the full protocol can be found at http://www.welfarequalitynetwork.net/. The protocol contains 33 measures which are measured or scored on-farm by means of behavioural observations, Qualitative Behavioural Assessment (QBA), a human approach test (avoidance distance at the feed rack; ADF), a management questionnaire, a resource checklist and clinical scoring (Table 2). The measures are mostly animal-based, with the exception of the resource-based measures regarding tethering, mutilations and access to water and pasture. For the QBA, after observing the herd for 20 min, the expressive quality of 20 terms was scored (Welfare Quality® 2009). These terms all have emotional connotations which directly reflect the animal's welfare (Algers *et al* 2009). The measures were carried out in the following order: ADF, QBA, behavioural observations, clinical scoring, resource checklist and management questionnaire.

The data on these measures are aggregated through several integration steps. During the first step, all measures are aggregated into scores for 12 criteria on a scale of 0–100, reflecting the worst and best possible score, respectively. For instance, the percentage of lame and severely lame cows and the percentage of cows with hairless patches and/or lesions is used to calculate the score for the criterion 'absence of injuries'. During the next integration step, the scores of the welfare principles ('good feeding', 'good housing', 'good health' and 'appropriate behaviour') are calculated on a scale from 0–100 (again reflecting the worst and best possible score, respectively) based on the criteria scores. For instance, the principle 'good feeding' is aggregated from the scores for the criteria 'absence of prolonged hunger' and 'absence of prolonged thirst'. The third and last integration step entails the calculation of an overall classification, which can be one of four categories: not classified, acceptable, enhanced or excellent. Farms are considered 'excellent' when they score more than 50 on all principles, with more than 75 on two. 'Enhanced' is assigned when a farm scores more than 20 on all principles and more than 50 on two, while it is considered 'acceptable' when the farm scores at least 10 on all principles and more than 15 on two. When a farm does not reach the standards for acceptable, it is considered 'not classified'.

This aggregation method implies that the overall welfare category cannot be calculated if data are lacking for one (or more) measure. Since some farms in the current study lacked data on individual somatic cell count, we assigned the same percentage to all farms in the study. That percentage was based on a sample of 121 Belgian dairy farms, where the average percentage of animals with a somatic cell count of > 400.000 was 13.8%.

Data analysis

The outcome of the WQ assessment was analysed at all levels (ie, at the level of the separate measures, the criteria, the principles, and the overall welfare categorisation). The non-binary outcomes at the first three levels of integration were modelled with linear mixed regression models with time (beginning or end of the indoor period) as fixed effect and farm as random intercept, in order to correct for the repeated measurements made on each farm. The analysed data were considered sufficiently normally distributed based on the graphical evaluation (histogram and QQ plot) of the residuals. All analyses were performed using proc mixed in SAS 9.4 (SAS Institute Inc, Cary, NC, USA). The overall classification 'acceptable' or 'enhanced', (as none of the herds in our sample turned out to be 'non-classified' or 'excellent') and other binary or proportion outcomes were analysed using a logistic mixed regression model (using Proc Genmod) similar to the previously described model.

Inter-observer reliability was calculated for the 20 WQ herd assessments using Intra-class Correlation Coefficients (ICC; following Shrout & Fleiss 1979) for all measures that were assessed separately but independently by the observers. This excludes measures for the criteria 'absence of prolonged thirst', 'absence of pain induced by management procedures' and 'expression of other behaviour', and some measures within 'absence of disease' (% of cows that suffered from dystocia, % of downer cows and % of mortality). Reliability was rated as 'moderate' for values between 0.41–0.60, as 'good' for values between 0.61–0.80 and as 'excellent' for values above 0.80 (following Landis & Koch 1977). For a small majority of measures, ICCs were either excellent (n = 7) or good (n = 5; Table 3). Insufficient ICCs were within the criteria 'absence of prolonged hunger' (% lean cows), 'comfort around resting' ('mean time to lie down' and '% cows colliding with housing equipment when lying down'), 'good human-animal relationship' ('% of cows that could be approached by 50 cm' and 'by 100 cm' and '% of cows that could not be approached') and within 'positive emotional state' ('Qualitative Behaviour Assessment Index').

Results

The percentage of cows with severe lameness, with mild integument alterations (hairless patches), and with vulvar discharge were lower at the beginning than the end of the indoor period (Table 4). Moreover, the number of coughs per cow per 15 min. and the mortality rate were lower, and the QBA-index higher (ie better), at the beginning than the end (Table 4). However, mean time to lie down and the percentage of cows with dirty udders were better at the end than the beginning. The other measures did not differ significantly between both assessment times (Table 4).

Farms received a better (ie higher) score for the principles 'good health' and 'appropriate behaviour' at the beginning than at the end of the indoor period (Table 5). Within 'good health', the criteria 'absence of injuries' and 'absence of

Table 3 Inter Correlation Coefficient (ICC) values between observers for those measure scores that were assessed separately but independently collected at ten farms at the beginning and end of the indoor period.

Criteria	Measures	ICC value
Absence of prolonged hunger	% lean cows	0.31
Comfort around resting	Mean time to lie down (s)	0.28
	% cows colliding with housing equipment when lying down	0.29
	% of cows lying outside lying area	0.66
	% of cows with dirty udder	0.85
	% of cows with dirty flank/upper leg	0.88
	% of cows with dirty lower legs	0.88
Absence of injuries	% of moderately lame cows	0.41
	% of severely lame cows	0.55
	% of cows with mild integument alterations	0.88
	% of cows with severe integument alterations	0.75
Absence of disease	Number of coughs per 15 min	0.72
	% of cows with nasal discharge	0.86
	% of cows with ocular discharge	0.90
	% of cows with hampered respiration	*
	% of cows with diarrhoea	0.93
	% of cows with vulvar discharge	0.61
Expression of social behaviour	Number of headbutts per cow per hour	0.68
	Number of other aggressive events per cow per hour	0.56
	% of cows that could be touched	0.42
Good human-animal relationship	% of cows that could be approached by 50 cm	0.14
	% of cows that could be approached by 100 cm	0.08
	% of cows that could not be approached	0.40
Positive emotional state	Qualitative Behaviour Assessment Index	0.13

*ICC could not be calculated.

disease' had a higher (ie better) score at the beginning than the end of the indoor period. Within 'appropriate behaviour', only the criterion 'positive emotional state', assessed with the QBA, was scored better at the beginning than the end of the indoor period.

The farm classification (across the four overall welfare categories) did not differ significantly when assessed at the beginning versus the end of the indoor period ($P = 0.580$). At the beginning of the indoor period, three farms were classified as acceptable (30%) and seven as enhanced (70%). At the end of the indoor period, four farms (the same as those in the beginning plus one) were classified as acceptable (40%) and six as enhanced (60%). None of the farms were categorised as not-classified or excellent.

Discussion

Although the overall categorisation was not significantly affected by time of assessment, the herds in this study obtained predominantly better scores for various welfare measures, criteria and principles when assessed at the beginning instead of the end of the indoor period. Given that our sample of farms was restricted both in terms of size and type and given that assessments were performed during a single indoor period only, caution is warranted when extrapolating these results to other herds of dairy cattle or other climatic conditions. Moreover, it cannot be excluded that differences reported in this study were due to other factors (such as a difference in outdoor temperature, or another seasonal effect) than the timing of the assessment

https://doi.org/10.7120/09627286.26.2.213 Published online by Cambridge University Press

Table 4 Median, minimum and maximum scores and *P*-values of the measure scores* collected at ten farms at the beginning and end of the indoor period.

Criteria	Measures	Beginning	End	*P*-value
Absence of prolonged hunger	% lean cows	4.6 (0.0–6.7)	6.4 (1.4–10.0)	0.678
Comfort around resting	Mean time to lie down (s)	6.0 (4.8–6.7)	5.0 (3.8–6.9)	0.011
	% cows colliding with housing equipment when lying down	19.6 (0.0–50.0)	30.4 (6.3–87.5)	0.110
	% of cows lying outside lying area	3.8 (0.0–22.5)	6.4 (0.0–19.1)	0.915
	% of cows with dirty udder	48.6 (4.6–81.8)	28.4 (3.0–63.6)	0.033
	% of cows with dirty flank/upper leg	50.0 (4.6–79.5)	63.6 (6.0–89.4)	0.225
	% of cows with dirty lower legs	89.7 (44.3–100.0)	90.8 (41.8–100.0)	0.457
Absence of injuries	% of moderately lame cows	12.8 (3.4–40.9)	25.6 (10.0–30.7)	0.479
	% of severely lame cows	2.2 (0.0–23.2)	7.3 (2.3–30.0)	0.032
	% of cows with mild integument alterations	19.8 (6.3–34.3)	52.2 (15.2–73.3)	< 0.001
	% of cows with severe integument alterations	19.6 (0.0–57.5)	22.8 (11.9–66.7)	0.408
Absence of disease	Number of coughs per 15 min	0.0 (0.0–0.1)	0.1 (0.0–0.2)	0.024
	% of cows with nasal discharge	17.3 (0.0–30.0)	16.9 (7.5–50.0)	0.126
	% of cows with ocular discharge	0.0 (0.0–18.2)	0.0 (0.0–14.0)	0.518
	% of cows with hampered respiration	0.0 (0.0–0.0)	0.0 (0.0–0.0)	–
	% of cows with diarrhoea	0.0 (0.0–10.6)	0.0 (0.0–18.6)	0.295
	% of cows with vulvar discharge	0.0 (0.0–3.0)	1.5 (0.0–3.3)	0.014
	% of cows that suffered from dystocia	9.2 (0.0–51.5)	15.9 (1.8–41.7)	0.337
	% of downer cows	3.3 (0.0–20.8)	3.4 (1.3–31.3)	0.939
	% of mortality	4.0 (2.0–7.5)	5.5 (2.1–13.0)	0.008
Expression of social behaviour	Number of head butts per cow per hour	0.3 (0.2–0.6)	0.4 (0.1–0.8)	0.426
	Number of other aggressive events per cow per hour	0.3 (0.1–0.7)	0.3 (0.0–0.6)	0.608
Good-human relationship	% of cows that could be touched	24.6 (2.9–43.2)	26.0 (10.6–45.9)	0.417
	% of cows that could be approached by 50 cm	37.4 (28.6–49.3)	34.1 (23.0–43.3)	0.472
	% of cows that could be approached by 100 cm	22.3 (10.2–36.7)	20.6 (10.6–30.3)	0.964
	% of cows that could not be approached	17.9 (3.0–34.3)	17.3 (4.6–31.3)	0.834
Positive emotional state	Qualitative Behaviour Assessment Index	1.8 (0.0–2.8)	0.6 (–3.1–1.4)	0.013

* Measures for 'absence of prolonged thirst', 'absence of pain induced by management procedures' and 'expression of other behaviour' are not shown as these resource- or management-based measures cannot differ over time when no significant changes have been made to the farm.

relative to the end of the pasturing period. Additionally, measures (especially ADF and QBA) with insufficient IOR were found, although this may partly be due to the low number of observers, or because samples of observed cows were not uniform for both assessor of each farm. Although previous studies were performed into the IOR of the various WQ measures for dairy cattle welfare, reliability issues require more thorough investigation (for a review, see Knierim & Winckler 2009). Further studies are needed into the IOR of some measures of WQ, and whether this could be improved, eg by means of training, sample selection or refinement of the measures. This may also warrant using a (control) group of farms which is observed twice but, eg during the same week, to examine any influence of test-retest reliability problems. For the current study, the low ICC for a considerable number of measures

Table 5 LS means (± SEM) and *P*-values for the principles and corresponding criteria scores during the beginning and end of the indoor period, higher scores indicate better welfare.

Principles	Criteria	Beginning	End	*P*-value
Good feeding		56.8 (6.6–100.0)	61.4 (10.9–91.8)	0.760
	Absence of prolonged hunger	69.5 (33.7–100.0)	59.8 (46.5–88.7)	0.422
	Absence of prolonged thirst	60.0 (3.0–100.0)	80.0 (3.0–100.0)	0.527
Good housing		54.7 (42.1–65.2)	53.6 (42.1–93.2)	0.460
	Comfort around resting	28.6 (8.7–45.2)	26.8 (8.7–89.3)	0.460
	Ease of movement*	100.0 (100.0–100.0)	100.0 (100.0–100.0)	–
Good health		35.5 (28.8–56.3)	30.4 (22.3–40.5)	0.016
	Absence of injuries	48.1 (29.0–90.7)	33.1 (17.3–52.2)	0.018
	Absence of diseases	36.9 (24.6–56.7)	30.2 (17.7–44.9)	0.024
	Absence of pain induced by management procedures*	28.0 (20.0–75.0)	28.0 (20.0–75.0)	–
Appropriate behaviour		52.7 (19.7–62.7)	49.8 (20.0–57.9)	0.015
	Social behaviour	77.0 (55.2–91.2)	76.0 (56.8–98.4)	0.0865
	Other behaviour*	79.8 (0.0–84.4)	79.8 (0.0–84.4)	–
	Human-animal relationship	35.7 (25.4–55.9)	36.4 (26.7–59.7)	0.826
	Positive emotional state	68.9 (50.0–78.5)	56.4 (25.5–65.2)	0.013

* Measures cannot differ over time when no significant changes have been made to the farm.

strengthens the choice for using two observers at the same time, to avoid influence of observer bias on the outcomes. As the step-wise integration method used by WQ is subject to some criticism (de Vries *et al* 2013; Heath *et al* 2014; Buijs *et al* 2016) we discuss differences between the beginning versus the end of the indoor period separately for the actual welfare measures and for the integrated scores.

Effect of timing during the indoor period on individual WQ measures

Of all the measures tested during the current study, 30% (8/27) differed significantly at the beginning versus the end of the indoor period. Six of these differences indicated better welfare at the beginning of the indoor period, and two indicated better welfare at the end of the indoor period. For one of the measures that scored worse at the beginning ('mean time to lie down'), ICC score was insufficient. This is not in accordance with previous IOR testing for 'mean time to lie down' which was considered acceptable (Algers *et al* 2009).

The percentage of cows with dirty udders was higher at the beginning than at the end of the indoor period. This finding is not in accordance with Burow *et al* (2013) and Corazzin *et al* (2010) who reported more dirty lower hind legs, but not more dirty udders during the indoor season than during summer grazing. While dirty lower legs indicate soiled alleyways and other walking areas, dirty udders are mostly caused by faeces in the lying area and transferred soil from the tail (Cook 2002). The udders of the cows in the current study may possibly be cleaner towards the end of the indoor period because the udders are usually shaved while they are kept indoors. As dirty udders have been found to be an important risk factor for mastitis (eg Bartlett *et al* 1992; de Pinho Manzi *et al* 2012), SCC could be higher during the outdoor period for these farms. As SCC was not recorded during the current study, this should be tested further.

The only other measure, besides udder cleanliness, that improved at the end of the indoor period was the time needed to lie down. Previous studies have found a negative relationship between a more comfortable surface of the lying area (concrete versus rubber or straw) and duration of the lying down movement (Krohn & Munksgaard 1993; Herlin 1997). A possible explanation for the finding in the current study that time needed to lie down was higher (worse) in the beginning than the end of the indoor period, is that cows needed to get used again to lying down in cubicles following the outdoor period.

The percentage of severely lame cows and of animals with hairless patches were higher at the end than at the beginning of the indoor period. The latter is in line with previous reports of fewer skin lesion problems during or shortly after the outdoor period (Rutherford *et al* 2008; Corrazin *et al* 2010; Burow *et al* 2013). This measure reflects the risk of cows colliding with structures in the housing system (eg cubicle separations, feeding rack), and the influence of the lying surface on the integument alterations of the cows (Lombard *et al* 2010; Andreasen & Forkman 2012; Brenninkmeyer *et al* 2013). Lameness data in the current

https://doi.org/10.7120/09627286.26.2.213 Published online by Cambridge University Press

study are in accordance with earlier studies that compared cows kept in loose-housed barns to cows kept on pasture for one production cycle (−40 to 305 days relative to calving; Olmos *et al* 2009a) and the beneficial effect of access to pasture on lameness (Hernandez-Menzo 2007; Rutherford *et al* 2009). The latter can be explained by grass being a more comfortable walking surface (Hernandez-Mendo *et al* 2007). Also, animals kept at pasture tend to move more (Hernandez-Mendo *et al* 2007), which is related to better claw health (Bielfeldt *et al* 2005).

The percentage of cows with vulvar discharge, number of coughs per cow per hour and mortality rate were all lower at the beginning than at the end of the indoor period. Regarding vulvar discharge, the results from the current study are in agreement with Bruun *et al* (2002) and Olmos *et al* (2009b), who found higher odds for reproductive disorders in zero-grazing herds. They suggest that pasturing is related to better general health (due to a lower bacterial count) and a better musculature (which makes calving easier). Corazzin *et al* (2010) also found a reduction in coughing when access to pasture was given. Possible explanations are accumulation of dust indoors and the temperature difference (colder at the end of the indoor period due to seasonal change). However, the biological relevance of the very small median difference of 0.1 coughs per cow per 15 min in the current study may be questioned.

The mortality rate was based on data from an entire year preceding each assessment, thus eliminating the possibility of seasonal differences (ie due to the time elapsed since last at pasture or seasonal climatic differences). Studies, such as USDA (2007), have found between-years fluctuations in dairy cattle mortality. In the current study, the higher mortality rate at the end of the indoor period could reflect such a fluctuation. However, because of a combination of the data collection method (estimates of the farmer, which were not verified) and the small mean difference in mortality rate between beginning and end of the indoor period (1.5%) combined with the small herd sizes in the study sample, we question the biological relevance in the current study.

The assessment of 'positive emotional state' of the herd was also better at the beginning than the end of the indoor period. This criterion is based on qualitative behaviour assessment (QBA). Given the subjective nature of this method, it could be particularly prone to observer bias (Tuyttens *et al* 2014), which was reflected in an insufficient IOR during the current study.

Effect of timing during the indoor period on the integrated WQ scores

The current study did not find a significant effect of timing of the assessments on the overall aggregated welfare score, while some studies that used a different welfare index than WQ did find such a difference (eg better welfare in summer in Burow *et al* 2013). The fact that there is no measure for 'thermal comfort', that we did not collect data on SCC and that the resource-based measures did not differ during the course of the current study could have limited the chances to result in a difference in overall welfare classification. Nonetheless, a number of integrated scores at the level of the welfare criteria ('absence of injuries', 'absence of diseases' and 'positive emotional state' and at the level of the principles ('good health' and 'appropriate behaviour') were better at the beginning than the end of the indoor period. However, the criterion 'comfort around resting' did not differ even though two measures did change. Recently, discussion has arisen about the integration method of the WQ protocol. Sensitivity analysis by de Vries *et al* (2013) showed that overall classification was very sensitive to certain measures (drinker space and collisions with housing equipment) while it was insensitive to most other measures they tested, even those consistently mentioned in literature as indicators of the biggest welfare problems (eg lameness and mortality). In line with this, Heath *et al* (2014) found that a single criterion, namely 'absence of prolonged thirst', correctly predicted the classification of 88% of the farms. These reports challenge whether the impact of the various welfare measures on overall welfare categorisation is a correct reflection of their relative importance rather than an unintended artefact of the complex step-wise integration procedure. For the current study, specifically, the lack of effect of measuring moment on overall welfare category likely has more to do with the integration method which is not sensitive enough to changes in individual scores, rather than the effect on overall welfare not being present.

Conclusion

Results from this study indicate that outcomes of animal-based evaluations of cattle that are kept under different circumstances during part of the year (eg with vs without access to pasture) may differ depending on the timing of the assessment. Nearly all differences indicated better animal welfare at the beginning of the indoor period. This indicates a positive effect of access to pasture in our sample of herds, and highlights the importance of the timing of animal-based welfare assessments for herds that are given access to pasture during summer, if results are to be comparable. Therefore, we recommend that timing of on-farm welfare assessments using the WQ protocol for dairy cattle should be standardised when comparisons between or within farms are made.

Acknowledgements

We thank Thijs Decroos for training the assessors and performing farm visits. Additionally, Erlijn Deolet, Emely Behaeghel and Annelies Coenraets, who performed the farm visits, are thanked. We are also grateful to the farmers for allowing us to come to their farms and perform the WQ protocol. Finally, we would like to thank Stephanie Buijs for proof-reading the manuscript.

References

Algers B, Bokkers EAM, Boivin X, Brörkens N, Canali E, Dalmau A, De Rosa G, Forkman B, Heutinck LFM, Keeling LJ, Knierim U, Laister S, Leach KA, Lensink BJ, Leruste H, Lolli S, MacKintosh N, Miljard F, Minero M, Napolitano F, Plesch G, Quast R, Regner AM, Chulze-Westerath H, Schmied C, van der Werf JTN, van Reenen CG, Velarde A, Waiblinger S, Wemelsfelder F, Westin R, Whay HR, Winckler C, Windschnurer I, Wolthuis-Fillerup M, Zegner K and Zucca D 2009 *Assessment of animal welfare measures for dairy cattle, beef bulls and veal calves.* Cardiff University: Cardiff, UK

Andreasen SN and Forkman B 2012 The welfare of dairy cows is improved in relation to cleanliness and integument alterations on the hocks and lameness when sand is used as stall surface. *Journal of Dairy Science 95*: 4961-4967. https://doi.org/10.3168/jds.2011-5169

Bartlett PC, Miller GY, Lance SE and Heider LE 1992 Environmental and managerial determinants of somatic cell counts and clinical mastitis incidence in Ohio dairy herds. *Preventive Veterinary Medicine 14*: 195-207. https://doi.org/10.1016/0167-5877(92)90016-9

Bielfeldt JC, Badertscher R, Tölle KH and Krieter J 2005 Risk factors influencing lameness and claw disorders in dairy cows. *Livestock Production Science 95*: 265-271. https://doi.org/10.1016/j.livprodsci.2004.12.005

Brenninkmeyer C, Dippel S, Brinkmann J, March S, Winckler C and Knierim U 2013 Hock lesion epidemiology in cubicle housed dairy cows across two breeds, farming systems and countries. *Preventive Veterinary Medicine 109*: 236-245. https://doi.org/10.1016/j.prevetmed.2012.10.014

Bruun J, Ersbøll AK and Alban L 2002 Risk factors for metritis in Danish dairy cows. *Preventive Veterinary Medicine 54*: 179-190. https://doi.org/10.1016/S0167-5877(02)00026-0

Buijs S, Ampe B and Tuyttens FAM 2016 Sensitivity of the Welfare Quality® broiler chicken protocol to differences between intensively reared indoor flocks: which factors explain overall classification? *Animal 15*: 1-10

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*: 834-842. https://doi.org/10.1017/S1751731112002297

Burow E, Thomsen PT, Sørensen JT and Rousing T 2011 The effect of grazing on cow mortality in Danish dairy herds. *Preventive Veterinary Medicine 100*: 237-241. https://doi.org/10.1016/j.prevetmed.2011.04.001

Cook NB 2002 The influence of barn design on dairy cow hygiene, lameness and udder health. *Proceedings of the 35th Annual Conference Proceedings of the American Association of Bovine Practitioners* pp 97-103. 26 September 2002, Rome, Italy

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*: e59. https://doi.org/10.4081/ijas.2010.e59

de Pinho Manzi M, Nóbrega DB, Faccioli PY, Troncarelli MZ, Menozzi BD and Langoni H 2012 Relationship between teat-end condition, udder cleanliness and bovine subclinical mastitis. *Research in Veterinary Science 93*: 430-434. https://doi.org/10.1016/j.rvsc.2011.05.010

de Vries M, Bokkers EAM, van Schaik G, Botreau RI, 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*: 6264-6273. https://doi.org/10.3168/jds.2012-6129

Federale Overheidsdienst Economie 2016 *Landbouwcijfers 2015*. http://statbel.fgov.be/nl/modules/publications/statistiques/economie/downloads/agriculture_-_chiffres_d_agricole_de_2015.jsp

Heath CAE, Browne WJ, Mullan S and Main DCJ 2014 Navigating the iceberg: reducing the number of parameters within the Welfare Quality® assessment protocol for dairy cows. *Animal 8*: 1978-1986. https://doi.org/10.1017/S1751731114002018

Herlin A 1997 Comparison of lying area surfaces for dairy cows by preference, hygiene and lying down behaviour. *Swedish Journal of Agricultural Resources 27*: 189-196

Hernandez-Mendo O, Von Keyserlingk MAG, Veira DM and Weary DM 2007 Effects of pasture on lameness in dairy cows. *Journal of Dairy Science 90*: 1209-1214. https://doi.org/10.3168/jds.S0022-0302(07)71608-9

Knierim U and 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*: 451-458

Krohn CC and Munksgaard L 1993 Behaviour of dairy cows kept in extensive (loose housing/pasture) or intensive (tie stall) environments II. Lying and lying-down behaviour. *Applied Animal Behaviour Science 37*: 1-16. https://doi.org/10.1016/0168-1591(93)90066-X

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

Lombard JE, Tucker CB, von Keyserlingk MAG, Kopral CA and Weary DM 2010 Associations between cow hygiene, hock injuries and free stall usage on US dairy farms. *Journal of Dairy Science 93*: 4668-4676. https://doi.org/10.3168/jds.2010-3225

Olmos G, Boyle L, Hanlon A, Patton J, Murphy JJ and Mee JF 2009a Hoof disorders, locomotion ability and lying times of cubicle-housed compared to pasture-based dairy cows. *Livestock Science 125*: 199-207. https://doi.org/10.1016/j.livsci.2009.04.009

Olmos G, Mee JF, Hanlon A, Patton J, Murphy JJ and Boyle L 2009b Peripartum health and welfare of Holstein-Friesian cows in a confinement-TMR system compared to a pasture-based system. *Animal Welfare 18*: 467-476

Onyiro OM and Brotherstone S 2008 Genetic analysis of locomotion and associated conformation traits of Holstein-Friesian dairy cows managed in different housing systems. *Journal of Dairy Science 91*: 322-328. https://doi.org/10.3168/jds.2007-0514

https://doi.org/10.7120/09627286.26.2.213 Published online by Cambridge University Press

Rutherford KM, Langford FM, Jack MC, Sherwood L, Lawrence AB and Haskell MJ 2009 Lameness prevalence and risk factors in organic and non-organic dairy herds in the United Kingdom. *The Veterinary Journal 180*: 95-105. https://doi.org/10.1016/j.tvjl.2008.03.015

Rutherford KMD, Langford FM, Jack MC, Sherwood L, Lawrence AB and Haskell MJ 2008 Hock injury prevalence and associated risk factors on organic and nonorganic dairy farms in the United Kingdom. *Journal of Dairy Science 91*: 2265-2274. https://doi.org/10.3168/jds.2007-0847

Shrout PE and Fleiss JL 1979 Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin 86*: 420-428. https://doi.org/10.1037/0033-2909.86.2.420

Tuyttens FAM, de Graaf S, Heerkens JL, Jacobs L, Nalon E, Ott S, Stadig L, Van Laer E and Ampe B 2014 Observer bias in animal behaviour research: can we believe what we score, if we score what we believe? *Animal Behaviour 90*: 273-280. https://doi.org/10.1016/j.anbehav.2014.02.007

USDA 2007 *Part II: changes in the US dairy cattle industry, 1991-2007*. USDA APHIS Veterinary Services Centers for Epidemiology and Animal Health, Fort Collins, Colorado, USA

Washburn SP, White SL, Green Jr JT and Benson GA 2002 Reproduction, mastitis, and body condition of seasonally calved Holstein and Jersey cows in confinement or pasture systems. *Journal of Dairy Science 85*: 105-111. https://doi.org/10.3168/jds.S0022-0302(02)74058-7

Welfare Quality® 2009 *Welfare Quality® Assessment Protocol for Cattle*. Welfare Quality® : Lelystad, The Netherlands