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

Animal Welfare 2014, 23: 81-94
ISSN 0962-7286
doi: 10.7120/09627286.23.1.081

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

SN Andreasen*, P Sandøe and B Forkman

Department of Large Animal Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, Grønnegårdsvej 8, 1870 Frederiksberg, Denmark
* Contact for correspondence and requests for reprints: sinen@sund.ku.dk

Abstract

Welfare assessment protocols using primarily animal-based measures are believed to give valid information about the welfare of animals. However, they tend to be time consuming and therefore costly, thus in practice there is often considerable reluctance to use them. In the present study, the relatively quick to use, simple but non-validated welfare assessment protocol for dairy cattle developed by the Danish Cattle Federation and the validated comprehensive Welfare Quality® protocol were compared in Danish conditions. In total, 44 Danish dairy herds were evaluated using the two protocols. The protocols were correlated on four sub-levels (corresponding to the 'Principles' in the Welfare Quality® protocol) and on the Overall welfare score. They correlated significantly with regard to the Principles 'Good health' and 'Appropriate behaviour'. Significant correlations were not found for 'Good feeding', 'Good housing', and the Overall score. On the basis of this we changed the Danish Cattle Federation protocol by introducing six new measures, changing three measuring procedures and omitting two measures. This extended protocol was found to correlate significantly with the Welfare Quality® protocol in all four Principles and on the Overall score. The extended protocol still has the advantage of the original Danish Cattle Federation protocol whereby under Danish field conditions it will take only 2 h to apply as opposed to 7–8 h for the Welfare Quality® protocol. We believe that the extended protocol balances, in a good way, the demands of practicability against its value as a diagnostic test.

Keywords: *animal-based measures, animal welfare, dairy cattle, on-farm assessment, welfare-assessment protocol, Welfare Quality®*

Introduction

Assessment of welfare at farm level seems to give rise to a dilemma between validity and cost. Assessment protocols using primarily animal-based measures can afford direct and valid information regarding how animals are affected by the way they are being kept and treated, but they tend to be time consuming and costly. In practice, therefore, there is often considerable reluctance to use them. Instead, welfare is often measured by simpler, but less valid, resource-based measures. The objective of this paper is to dispel this dilemma and ascertain whether it is possible to find an animal-based welfare protocol for dairy cattle which is at the same time speedy, cost effective and measures states which, collectively, provide a valid picture of the welfare of the assessed animals.

A wide range of welfare assessment protocols for dairy cattle currently exist, and new ones are under development. The number and nature of measures included vary widely. However, measures included in the protocols are either resource- /management-based, animal-based or both (Rushen & de Passillé 1992; Waiblinger *et al* 2001). The goal of all these protocols is to establish welfare status. At present, animal-based measures are being emphasised in more and more protocols as it is believed that these measures get closer to the actual welfare enjoyed by the animal (Johnsen *et al* 2001; Blokhuis *et al* 2010).

One of the most comprehensive and, at the same time, validated welfare assessment tools are the Welfare Quality® protocols (WQ). The WQ programme was designed to develop European standards for on-farm welfare assessment (Blokhuis *et al* 2013). Its main underlying idea is that welfare is determined with reference to what matters to the affected animals, ie their mental states (Botreau *et al* 2007). Because of this, the WQ protocols primarily involve animal-based measures of states that are assumed to be good indicators of underlying mental states. However, due both to limitations in what is possible to measure and stakeholder concerns, the protocol contains some resource-based measures (Miele *et al* 2011). The protocol concerning dairy cattle includes 29 validated measures, and of these only nine are resource- /management-based (Forkman & Keeling 2009).

The WQ protocols are integrated in 12 criteria, four Principles and one Overall score. The integration procedure is complex and based on judgments made by a number of experts; this approach has resulted in a procedure where different measures are given different amounts of emphasis (Botreau *et al* 2009). The scores assigned to the farms are not relativised to the population; rather an absolute scale is used. Carrying out the protocol for dairy cattle is time consuming: on a farm with 200 cows it takes approximately 7 h 45 min to conduct (Welfare Quality® 2009).

This time requirement and the associated costs are seen by many as a problem. Because of this there is a case for simplifying the protocols. However, so far, this has proved a difficult task. A recent study attempted to evaluate a reduced WQ protocol whereby a set of WQ measures were replaced by predictions building upon the remaining measures; the aim was to see whether on-farm assessment time could be reduced. Agreement among the observed and predicted values was poor to moderate, however, and ultimately the modified approach was not recommended (de Vries *et al* 2013).

As in many other European countries, levels of public awareness of animal welfare issues are high in Denmark, and in 2005/2006 the Danish Cattle Federation (DCF) decided to include animal welfare in their business guideline. They developed a dairy cow welfare assessment protocol using the hedonistic view that (animal) welfare is constituted by both the positive and negative experiences, similar to the definition used by WQ. The protocol emphasises that the information needed has to be obtained in the shortest time possible using scientific and practical knowledge (Danish Cattle Federation 2005). The DCF protocol contains ten animal-based measures, and can be completed in approximately 2 h on a farm with 200 dairy cows. The DCF protocol is used as one basis of the present study.

The protocol was intended to help the farmer achieve relative improvements in the welfare of his or her livestock. Since it was not originally intended to be used in quality assurance, or for purposes of certification, a method for summing up the protocol using absolute values has not yet been established.

Part of the purpose of the animal welfare initiatives set up by the DCF was to engage in a dialogue with society over concerns about animal welfare. Ingemann *et al* (2009) studied the protocol from this perspective. It was concluded that the measures used in the DCF protocol might pose a challenge. The authors claimed that the measures used could be associated more with economic benefits and acceptance by farmers than with the need to cover all relevant aspects of animal welfare, and it was suggested that this possible conflict could contribute to problems in the public dialogue and to biases in the protocol. To check for such biases it is helpful to compare the existing DCF protocol with the more comprehensive WQ protocol.

The objective of this study was to ascertain whether or not correlation could be found between the comprehensive WQ protocol and the existing, less time-consuming DCF protocol in Danish conditions (Part 1). The DCF protocol is not as all-embracing as the WQ protocol. Therefore, it can be assumed that it may be necessary to extend the DCF protocol. If the need for extension is found in our results from correlating the original DCF protocol and the WQ protocol, it will be further investigated whether extension of the DCF protocol will ameliorate correlation (Part 2). If the DCF protocol, or a slightly modified version of it, can deliver results that are in accordance with the Welfare Quality® protocol it will be shown that it is possible to simplify animal-based welfare assessment, thereby saving time and making animal-based welfare assessment more practicable — at least in Danish conditions.

Part 1 — A comparison of Welfare Quality® and Danish Cattle Federation protocols of animal welfare

Materials and methods

The study was performed as a cross-sectional observational study. Data were collected during the autumn and winter of 2010/2011 (October–March) in Denmark. Farms in the study were evaluated using the two welfare assessment protocols: the Welfare Quality® protocol (WQ) and the existing protocol designed by the DCF. All data were collected by the same observer (SNA), who had received the necessary training in the Welfare Quality® protocol.

Study herds

In total, 63 farms were invited to participate in the study. The farms had more than 50 cows, all of which were Danish Holstein-Friesian and used loose housing. Of those invited, 44 agreed to participate.

The total number of cows was 8,106, with a mean of 184 cows on each farm (range: 101–452). Two of the farms were organic. Overall, four used deep bedding (the rest using cubicles), 17 used automatic milking, and 27 milked in parlour.

During data collection, all dairy cows were in the facility. Primiparous, multiparous and dry cows were included, whereas cows housed in sick pens, young stock and calves were excluded.

Welfare assessment protocols

The Welfare Quality® protocol

A welfare assessment using the WQ protocol (Welfare Quality® 2009) was performed on each farm. This protocol consists of 29 primarily animal-based measures (Forkman & Keeling 2009), which are used to calculate the satisfaction of 12 Criteria (C1–C12). The 12 Criteria are aggregated into four Principles (P1–P4) which are ultimately translated into an Overall score (Overall; OV). The farms obtain a score for each measure, one for each Criterion, one for each Principle, and in the end an Overall score. On the three latter levels the farms can be placed into one of four welfare categories: 'Not classified', 'Acceptable', 'Enhanced' or 'Excellent'. The welfare categories assigned in the Principles and in the Overall score are the ones most used. The score assigned to the farm is not relativised to the welfare status of the population; in other words, scores are absolute (Table 1 summarises the WQ protocol, for further description of the measures see Welfare Quality® 2009).

https://doi.org/10.7120/09627286.23.1.081 Published online by Cambridge University Press

Table 1 Summary of the Welfare Quality® (WQ) and the Danish Cattle Federation (DCF) protocols (original and extended).

		WQ		DCF
P1	*Principle*†	*Criteria*‡	*Measures*	*Measures*
	Good feeding	Absence of prolonged hunger (C1)	Body Condition Score (BCS)	Body Condition Score (BCS): *cows were assigned score 1–5. Summed up in three levels, BCS ≤ 2, 2 < BCS < 4 and BCS ≥ 4*
		Absence of prolonged thirst (C2)	'Water provision'	No measures
			'Cleanliness of water points'	
			'Water flow'	
			'Functioning of water points'	
P2	Good housing	Comfort around resting (C3)	'Time needed to lie down'	Getting-up behaviour: *cows rising were evaluated on 5 levels, 1–2 untroubled getting up, 3–5 troubled getting up*
			'Animals colliding with housing equipment during lying down'	
			'Animals lying partly or completely outside the lying area'	Cleanliness (udder, thigh and legs): *cows were assessed from behind and from one side on four levels, 1–2 clean, 3–4 dirty*
			'Cleanliness of udder/upper legs and lower legs'	
				Animals lying partly or completely outside the lying area
		Thermal comfort (C4)	No measures	No measures
		Ease of movement (C5)	'Presence of tethering'	No measures
			'Access to outdoor loafing area or pasture'	
P3	Good health	Absence of injuries (C6)	'Lameness'	Lameness: *cows were assessed standing and walking on four levels, 0 no lameness, 2, 4 and 2G is lameness*
			'Integument alterations'	
				Integument alterations: *cows are assessed from one side on four levels, 1 no alterations, 2 early stage alterations, 3–4 alterations*
		Absence of disease (C7)	'Coughing'	Coat condition: *the coat was assessed using four levels, sleek, dull, hair loss and ringworm*
			'Nasal discharge'	
			'Ocular discharge'	Hoof condition: *hoofs were assessed using three levels, no remarks, length and asymmetry which were summed up to remarks*
			'Hampered respiration'	
			'Diarrhoea'	
			'Vulvar discharge'	
			'Milk somatic cell count'	
			'Mortality'	
			'Dystocia'	
			'Downer cows'	
		Absence of pain by management procedures (C8)	'Disbudding/dehorning'	No measures
			'Tail docking'	
P4	Appropriate behaviour	Expression of social behaviours (C9)	'Agonistic behaviour'	Positive behaviour: *presence of grooming (social licking) and caudal licking*
		Exp of other behaviours (C10)	'Access to pasture'	No measures
		Good human-animal relationship (C11)	'Avoidance distance'	Confidence: *measured as avoidance distance in the barn on five levels; 1–3 describes sufficient confident behaviour, 4–5 describes insecure behaviour*
		Positive emotional state (C12)	'Qualitative Behaviour Assessment'	No measures

The DCF protocol is described in detail in Appendix 1. For a more detailed description of the WQ protocol, see Welfare Quality (2009).
† P1–P4: Principle 1 to 4; ‡ C1–C12: Criteria 1 to 12.

Table 2 Presentation of the percentiles (rounded figures in percent) found for the results for the 44 farms when using the Danish Cattle Federation protocol (see Appendix 1).

Factor	25% best farms (Score 4)	25% second-best farms (Score 3)	40% second-worst farms (Score 2)	10% worst farms (Score 1)
Untroubled getting up	100	94–99	79–93	63–78
Sufficient confident behaviour	95–100	88–94	73–87	51–72
Hoofs, no remarks	100	100	96–99	65–95
No lameness	87–100	77–86	55–76	44–54
Clean	89–100	74–88	44–73	26–43
No integument alterations	45–86	28–44	4–27	0–3
Early stage integument alterations	14–32	33–39	40–63	64–78
Integument alterations	0–14	18–27	28–49	50–67
BCS $\leq$ 2	0	0	1–8	9–14
2 < BCS < 4	100	97–99	87–96	82–86
BCS $\geq$ 4	0	0	1–8	9–12
Coat, sleek	100	100	97–99	93–96
Coat, dull			No animals	
Coat, hair-loss	0	0	0	1–4
Coat, ringworm	0	0	1–3	4–7

The percentiles are used to assign each farm a score for being either among the 25% best, the 25% second-best, the 40% second-worst or the 10% worst. For example, for a farm to receive score 4 (being among the 25% best) in 'Untroubled getting up' 100% of the animals must rise without problems, if only 80% of the animals rise without problems then the farm receives a score 2.

Table 3 Correlations between the Danish Cattle Federation protocol (original and extended) and the Welfare Quality® protocol.

	WQ and DCF protocol		WQ and extended DCF protocol	
	Spearman ρ	*P*-value	Spearman ρ	*P*-value
P1 Good feeding	–0.1894	0.2181	0.3937	0.0082*
P2 Good housing	0.1834	0.2333	0.6069	< 0.0001*
P3 Good health	0.6662	< 0.0001*	0.7785	< 0.0001*
P4 Appropriate behaviour	0.4541	0.0020*	0.3601	< 0.0163*
Overall score	0.1590	0.3026	0.4730	0.0012*

WQ: Welfare Quality® protocol;
DCF: Danish Cattle Federation protocol;
* Significant finding.

On-farm assessments started approximately 15 min after morning feeding (0415–0900h). The recommended sample size was used (Welfare Quality® 2009), in total 3,051 animals were inspected (mean 69 cows per farm, range: 51–91).

The existing welfare assessment protocol developed by the Danish Cattle Federation

The DCF welfare assessment protocol consists of ten animal-based measures: Lying-behaviour (ie the number of cows lying outside the cubicles were counted when the area of the cows was entered in the morning); Positive behaviour (ie grooming [social licking] and caudal licking); Getting-up behaviour; Confidence (measured by avoidance distance); Hoof condition; Integument alterations; Lameness; Body condition; Cleanliness of udder (including teats), thigh and legs; and Coat condition (see Table 1; for further description of the measures see Appendix 1 [available at the supplementary material to papers published in *Animal Welfare* section at the UFAW website, www.ufaw.org.uk]). During its development, the protocol was tested on 40 farms in Denmark and 60 in Sweden, but has never been released as a finished tool for use in Danish dairy production. In Sweden, a protocol resembling the DCF's is used by the Swedish Dairy Association (Svensk Mjölk, Fråga Kon) (Anonymous 2013a).

https://doi.org/10.7120/09627286.23.1.081 Published online by Cambridge University Press

Table 4 Water provision in the seven farms which ranked better in the DCF protocol.

Farm	Number of animals in group	Trough/bowl	Number of animals allowed if partly sufficient	Number of animals allowed if sufficient	Score WQ: 2, insufficient; 1, partly sufficient; 0, sufficient
1	117	2 × 125 cm	62	41	2
2	101	3 × 130 cm	97	65	2
	38	3 × 39 cm	29	19	2
3	133	3 × 170 cm	127	82	2
4	109	3 × 132 cm	99	66	2
5	172	5 × 110 cm	170	113	2
		1 × 68 cm			
6	90	1 × 115 cm	180	120	0
		2 × 270 cm			
		1 × 68 cm			
	61	2 × 119 cm	59	39	2
7	14	2 bowls	30	20	0
	105	2 × 183 cm	164	109	0
		1 × 290 cm			
	64	2 bowls	75	50	1
		1 × 183 cm			

Insufficient water provision: farm 1–6, partly insufficient water provision: farm 7.

In this study, the DCF protocol was performed together with, and on the same animals as, the WQ protocol. The recommended sample size was used; in total, 1,288 animals were inspected (mean 29 cows per farm, range: 26–40).

The measures used in the DCF protocol have never been put through calculations giving sub-scores or a final score. To be able to correlate the DCF protocol with the WQ protocol it was necessary to sum up the DCF measures. In the protocol used by the Swedish Dairy Association, measures are summarised using four percentiles: these place the farm among the 25% best farms, the 25% second-best farms, the 40% second-worst farms and the 10% worst farms for each measure. This approach was applied to the DCF protocol, and the farms were subsequently assigned a score for each measure — ie they received a score of 4 when they were among the 25% best and score of 1 when they were among the 10% worst. In order to approximate to the WQ protocol the measures were later summed up in relation to the four Principles and to the Overall score. In this procedure (in contrast with what happens in the WQ protocol) no weighting of the measures occurred (Table 1 summarises the DCF protocol).

Statistical analysis

For statistical analysis the software SAS JMP© 10 (SAS Institute Inc, SAS Campus Drive, Cary, NC, USA) was used.

Welfare Quality® protocol

The Welfare Quality® calculations, which were performed by INRA (Institut National de la Recherche Agronomique), France, followed the outline in the protocol (Welfare Quality® 2009). The data collected on farms were typed into Excel® spreadsheets, with one for each farm. Each dataset was randomly checked for entry errors.

The Danish Cattle Federation protocol

Data collected on the farms were typed into an Excel® spreadsheet for each farm and randomly checked for entry errors. Distributions for each measure were calculated and each farm assigned to one of the four percentiles, when among the 25% best they received score 4, when among the 10% worst they received score 1 etc (for percentiles, see Table 2). The measures for 'Lying behaviour' and 'Positive behaviour' were removed from the calculations, because no animals were found to be lying outside the cubicles and because grooming (social licking) and caudal licking occurred on all farms.

Comparison of the two welfare assessment protocols

The WQ protocol and the DCF protocol were compared using Spearman rank correlation. The correlations were performed on the Principles (P1–P4) and on the Overall score (OV). Correlation was performed on the continuous scores which were given to the farms and not on the welfare

Table 5 Data on the two Welfare Quality® (WQ) measures, 'Time needed to lie down' and 'Collisions with housing equipment during lying down'.

Farm	Mean time needed to lie down (s)	Collision with equipment (%)
Farms ranking better in the DCF protocol		
1	6.58	50.00
2	6.45	62.50
3	6.33	38.46
4	7.21	56.25
5	6.66	68.75
6	6.43	69.25
Farms ranking better in the WQ protocol		
1	6.25	10.00
2	5.99	6.67
3	6.72	16.67
4	4.97	54.55

DCF protocol: Danish Cattle Federation protocol.

categories (Excellent, Enhanced, Acceptable and Not classified). Where a Principle failed to attain significant correlation, the farms which differed by more than 20 Spearman ranks were identified and the measures contributing to that Principle tested. The goodness-of-fit test was used to determine whether or not data followed a normal distribution, and then either the Student's *t*-test or the Wilcoxon-Mann-Whitney test was performed to see if the measures in the two protocols agreed.

Results

Protocol correlation

Using the Spearman rank correlation a significant positive correlation was found between WQ P3 and DCF P3 ('Good health'), $\rho = 0.6818$, $P < 0.0001$ and between WQ P4 and DCF P4 ('Appropriate behaviour'), $\rho = 0.4541$, $P = 0.0020$. However, no significant correlation was found between WQ P1 and DCF P1 ('Good feeding'), WQ P2 and DCF P2 ('Good housing') and WQ OV and DCF OV (Overall assessment; for correlations; see Table 3).

In WQ P1, only lean animals are given a weight when the measures are summed up into Criteria and Principles (Welfare Quality® 2009). In DCF P1, all three levels of Body Condition Score (BCS) are assigned a score. To align the DCF and WQ protocols when looking at BCS, the three BCS levels were reduced to one (BCS $\leq$ 2). Following this correction, the correlation between WQ P1 and DCF P1 was still not significant ($\rho = -0.0769$, $P = 0.6198$), for that reason the three BCS levels used in the DCF protocol were retained.

Principle 1 — Good feeding

No significant correlation was found between WQ P1 and DCF P1. It was found that eight farms were ranked much higher in the DCF protocol and another eight farms were ranked much higher in the WQ protocol. The variations could be due to one or more of the 12 parameters included in Principle 1 (see Table 1): number of animals on farm; sample size; Criterion 1 (WQ); Criterion 2 (WQ); proportion of lean cows in WQ; proportion of fat cows in WQ; proportion of cows with BCS $\leq$ 2 in DCF; proportion of cows with 2 < BCS < 4 in DCF; proportion of cows with BCS $\geq$ 4 in DCF; water provision (sufficient, partly sufficient, insufficient) (WQ); cleanliness of water points (clean, partly clean, dirty) (WQ); water flow (sufficient, insufficient) (WQ); and functioning of water points (working correctly, malfunctioning) (WQ).

The eight farms which received a much better score in the DCF protocol had a significantly lower score in Criterion 2 (WQ), $P = 0.0063$. None of the farms had animals with a BCS $\geq$ 4 (DCF), $P = 0.0164$, and in seven out of the eight farms water provision was found to be either partly sufficient (one farm) or insufficient (six farms), $P = 0.0056$ (see Table 6). In the WQ protocol, sufficient water provision is defined as at least one water bowl for 10 cows or 6 cm of trough per cow; and partly sufficient water provision is defined as at least one water bowl for 15 cows or 4 cm trough per cow (Welfare Quality® 2009). Water provision is not measured in the DCF protocol (for water provision on the seven farms, see Table 4).

The eight farms which received a much better score in the WQ protocol had significantly more animals with BCS $\geq$ 4 (DCF), $P = 0.0032$ (see Table 6).

Principle 2 — Good housing

No significant correlation was found between WQ P2 and DCF P2. Six farms ranked much better in the DCF protocol, and another four ranked much better in the WQ protocol. The variation could be due to one or more of the 13 parameters included in Principle 2 (see Table 1): number of animals on farm; sample size; Criterion 3 (WQ); Criterion 5 (WQ); access to pasture/outdoor loaf; proportion of dirty cows (WQ); proportion of clean cows (WQ); proportion of cows colliding with equipment (WQ); proportion of cows lying outside the cubicles (WQ); proportion of clean cows (DCF); proportion of dirty cows (DCF); proportion of cows with normal getting-up behaviour (DCF); and proportion of cows with abnormal getting-up behaviour (DCF).

The six farms which received a much better score in the DCF protocol had a significantly lower score in Criterion 3 (WQ), $P = 0.0064$. These six farms had significantly more animals demonstrating normal getting-up behaviour (DCF), $P = 0.0292$ as well as significantly fewer animals showing abnormal getting-up behaviour (DCF), $P = 0.0292$ (see Table 6).

Criterion 3 in the WQ protocol covers behaviour around lying down. On the 44 farms, 560 lying events were recorded with a mean of 12.7 events on each farm (range:

https://doi.org/10.7120/09627286.23.1.081 Published online by Cambridge University Press

Table 6 Diverging farms are farms with a rank discrepancy of more than 20 between the results of the WQ protocol and those of the Danish Cattle Federation protocol. Measures included in the table are those that differ significantly between the diverging farms and all farms.

Principle 1 — Good feeding								
Farms ranking better in DCF protocol								
		Farms (n)	**Mean**	**SD**	**Lower 95%**	**Upper 95%**	***P*-value**	**Goodness-of-fit**
WQ C2	All	44	68.34	37.78	56.85	79.83	0.0063	< 0.0001
	Diverging farms	8	22.25	37.21	–8.86	53.36		
BCS ≥ 4 (DCF)	All	44	2.85	3.60	1.76	3.95	0.0164	< 0.0001
	Diverging farms	8	0.00	0.00	0.00	0.00		
Water provision (WQ)	All	44	0.68	0.80	0.44	0.93	0.0056	< 0.0001
	Diverging farms	8	1.63	0.74	1.00	2.25		
Farms ranking better in WQ protocol								
BCS ≥ 4 (DCF)	All	44	2.85	3.60	1.76	3.95	0.0032	< 0.0001
	Diverging farms	8	7.18	2.93	4.72	9.63		
Principle 2 — Good housing								
Farms ranking better in DCF protocol								
WQ C3	All	44	28.01	11.71	24.45	31.56	0.0064	< 0.0001
	Diverging farms	6	16.50	0.00	16.50	16.50		
% Normal getting up (DCF)	All	44	91.99	9.06	89.23	94.74	0.0292	< 0.0001
	Diverging farms	6	98.81	2.91	95.75	101.87		
% Abnormal getting up (DCF)	All	44	8.02	9.06	5.26	10.77	0.0292	< 0.0001
	Diverging	6	1.19	2.91	–1.87	4.25		
Farms ranking better in WQ protocol								
WQ C3	All	44	28.00	11.71	24.45	31.56	0.0348	< 0.0001
	Diverging farms	4	40.15	5.77	30.97	49.34		

DCF: Danish Cattle Federation protocol;
WQ: Welfare Quality® protocol;
BCS: Body Condition Score.

6–24). Time needed to lie down was recorded using a stopwatch. The mean length of time needed to lie down (farm level) was 6.01 s (range: 3.63–9.26 s). During the lying-down event it was noted whether the cow collided with the housing equipment. In WQ, a farm is considered to have a serious problem if lying down takes more than 6.30 s; a farm has a moderate problem if lying down takes between 5.20–6.30 s; and a farm is considered normal if lying down takes less than 5.20 s. A farm is likewise considered to have a serious problem when collisions with equipment are observed in more than 30% of the lying-down events; a farm has a moderate problem if collisions are seen in 20–30% of the lying-down events; and is normal if collisions occur in less than 20% of the lying-down events (Welfare Quality® 2009). Here, seven of the farms scored normally in terms of time needed to lie down, 18 had a moderate problem, and 19 had a serious problem. Where collisions were concerned, nine of the farms scored normally, three had a moderate problem, and 32 had a serious problem. For lying-down behaviour in the ten farms with a rank discrepancy of more than 20 ranks (see Table 5).

It was found that the four farms which received a much better score in the WQ protocol had a significantly higher score in Criterion 3 (WQ), $P = 0.0348$ (for results, see Table 6).

Discussion

A significant correlation was found between the WQ protocol and the DCF protocol in respect of two of the four Principles: P3 (Good health) and P4 (Appropriate behaviour). However, no such correlation was found for P1 (Good feeding) and P2 (Good housing).

Principle 1 — Good Feeding

Principle 1 contains measures for 'Absence of prolonged hunger' and 'Absence of prolonged thirst'. The score for 'Absence of prolonged hunger' in the WQ protocol uses a Body Condition Score (BCS) which categorises the cows as being lean, regular or fat. In the DCF protocol a 5-scale BCS is used, but the result is summarised to three levels (see Appendix 1). Examination of the three different BCS categories revealed significant agreement between the two protocols with respect to the scoring of lean animals, but not for animals scored as regular or fat. In the score generation procedure for the criterion 'Absence of prolonged hunger' in the WQ protocol only scores for animals which are lean are used (Welfare Quality® 2009). To render the results more comparable a decision was made to apply a score to the farms only for their percentage of animals with BCS ≤ 2 in the DCF protocol. This did not, however, result in a significant correlation between the two P1 scores. In P1 (see Table 6) the eight farms which ranked better in the DCF protocol had no animals with BCS ≥ 4. These farms were, as a consequence, placed among the 25% best farms and received a score of 4 in this measure. Given this high score, the farms will be interpreted as having better welfare in the DCF protocol. The eight farms ranking better in the WQ protocol had animals with BCS ≥ 4, resulting in a lower DCF score and an interpretation of poorer welfare.

The DCF protocol contains no measures for absence of prolonged thirst. The WQ protocol, on the other hand, contains four measures; these are all resource-based, as no useful animal-based measure was found (Forkman & Keeling 2009) (see Table 1). The 'Water provision' measure from the WQ protocol was found to affect the difference in ranking between the protocols. Of the eight farms which ranked better in the DCF protocol, seven had either only partly sufficient (one farm) or insufficient (six farms) water provision in the WQ protocol (for water provision, see Table 4); and because of poor water provision, the farms are scored as having poorer welfare in the WQ protocol than they have in the DCF protocol. Investigation of the seven farms in question showed that none of the animals from the sample were clinically dehydrated, as shown by lack of eyeball recession and assessment of skin elasticity. No queuing around the water points was observed on any of the farms, the water points were all functioning and all farms had at least one water provision site that was clean (as defined by the WQ protocol). Dairy cows thrive in temperatures between 5 to 25°C (Hemsworth *et al* 1995; Roenfeldt 1998). According to the Danish Meteorological Institute, mean summer-time temperatures in Denmark, over the period 2002–2011, were: June 14.8°C, July 17.3°C, and August 17.1°C (Anonymous 2013b). This could suggest that being unable to consume large quantities of water is less of a problem in Denmark since the mild climate would mean less water consumption than in a hotter climate. Access to water is, however, an important prerequisite of animal welfare (Hemsworth *et al* 1995; Igbokwe 1997; Farm Animal Welfare Committee 2013), and we believe that a measure for it should be included in the DCF protocol; moreover, this can be done without the protocol taking longer to perform.

It could be argued that the BCS scale used in the DCF protocol should be replaced with the BCS scale used in the WQ protocol, as the latter is validated and relatively simple (Leach *et al* 2009a). Fat cows risk developing metabolic disorders, and this could have implications for welfare (Roche *et al* 2009). Given this, it may be argued that the proportion of fat cows should be given negative weight in a welfare protocol. Again, this change would not affect the time taken to conduct the DCF protocol.

Principle 2 — Good housing

Principle 2 contains measures for 'Comfort around resting' (C3) and 'Ease of movement' (C5). In the WQ protocol, four measures are included in the criterion for 'Comfort around resting'; in the DCF protocol only two measures are included. In the WQ protocol two measures are included in the criterion for 'Ease of movement'; whereas no measures are included in the DCF protocol (see Table 1).

For this principle, the six farms which ranked better in the DCF protocol had significantly more animals with untroubled getting-up behaviour (DCF) and significantly fewer animals with troubled getting-up behaviour (DCF) (see Table 6). This gives the six farms a high (good) score in both getting-up behaviour categories (DCF). The six farms also had a significantly lower score for Criterion 3 ('Comfort around resting') in the WQ protocol, and thus better welfare according to the DCF protocol. The four farms ranking better in the WQ protocol had a significantly higher C3 score in the WQ protocol. As getting-up behaviour (DCF) is part of C3, and as no differences between the farms were found in C5 ('Ease of movement'), all differences in the principle 'Good housing' can be assigned to C3. The other measures included in this criterion are: cleanliness (WQ and DCF); time needed to lie down (WQ); animals colliding with housing equipment during lying down (WQ); and animals lying partly or completely outside the lying area (WQ). In terms of cleanliness, the two protocols agree significantly about the total number of dirty animals.

The DCF protocol includes animals lying partly or completely outside the lying area, but this was left out of the calculations because no animals were observed lying outside the cubicles when this was checked at the beginning of the DCF protocol (see Appendix 1). During data collection for the WQ protocol only two animals in all were scored as lying partly or completely outside the lying area, and therefore this measure could not be ascribed any significance. The six farms which ranked better in the DCF protocol were all scored as having a serious problem in 'Time needed to lie down' (WQ) and in 'Animals colliding with housing equipment during lying down' (WQ). The four farms which ranked better in the WQ protocol all scored normal on 'Time needed to lie down', and three of the four farms scored normal on collision with equipment (for behaviour around lying down, see Table 5).

https://doi.org/10.7120/09627286.23.1.081 Published online by Cambridge University Press

Table 7 Summary of the Welfare Quality® (WQ) and the extended Danish Cattle Federation (DCF) protocols. The DCF protocol is described in detail in Appendix I. For a more detailed description of the WQ protocol, see Welfare Quality® (2009).

	WQ		Extended DCF
Principle[†]	*Criteria*[‡]	*Measures*	*Measures*
Good feeding	Absence of prolonged hunger (C1)	Body Condition Score (BCS)	Body Condition Score (BCS): *Assessing procedure as the WQ protocol*
	Absence of prolonged thirst (C2)	'Water provision'	'Water provision'
		'Cleanliness of water points'	
		'Water flow'	
		'Functioning of water points'	
Good housing	Comfort around resting (C3)	'Time needed to lie down'	Cleanliness (udder. thigh and legs): *cows were assessed from behind and from one side on four levels, 1–2 clean, 3–4 dirty*
		'Animals colliding with housing equipment during lying down'	
		'Animals lying partly or completely outside the lying area'	Animals lying partly or completely outside the lying area
		'Cleanliness of udder, flank/upper legs and lower legs'	'Time needed to lie down' *Assessing procedure as the WQ protocol*
			Animals colliding with housing equipment while lying down *Assessing procedure as the WQ protocol*
	Thermal comfort (C4)	No measures	No measures
	Ease of movement (C5)	'Presence of tethering'	No measures
		'Access to outdoor loafing area or pasture'	
Good health	Absence of injuries (C6)	'Lameness'	Lameness: *assessing procedure as the WQ protocol*
		'Integument alterations'	Integument alterations: *cows are assessed from one side on four levels, 1 no alterations, 2 early stage alterations, 3–4 alterations*
	Absence of disease (C7)	'Coughing'	Coat condition: *the coat was assessed using four levels: sleek, dull, hair loss and ringworm*
		'Nasal discharge'	
		'Ocular discharge'	Hoof condition: *hoofs were assessed using: no remarks, length and symmetry*
		'Hampered respiration'	
		'Diarrhoea'	Dystocia: *assessing procedure as the WQ protocol*
		'Vulvar discharge'	Mortality: *assessing procedures as the WQ protocol*
		'Milk somatic cell count'	
		'Mortality'	Milk somatic cell count: *assessing procedure as the WQ protocol*
		'Dystocia'	
		'Downer cows'	
	Absence of pain by management procedures (C8)	'Disbudding/dehorning'	No measures
		'Tail docking'	
Appropriate behaviour	Expression of social behaviours (C9)	'Agonistic behaviour'	No measures
	Expression of other behaviours (C10)	'Access to pasture'	No measures
	Good human-animal relationship (C11)	'Avoidance distance'	Confidence: *assessing procedure as the WQ protocol*
	Positive emotional state (C12)	'Qualitative Behaviour Assessment'	No measures

[†] P1-P4: Principle 1 to 4;
[‡] C1-C12: Criteria 1 to 12.

From the above it can be concluded that differences between the two protocols' scores for the principle 'Good housing' are due to differences in the way behaviour related to resting is measured; and that the protocols agree on the scores for cleanliness. Getting-up behaviour from the DCF protocol cannot be directly compared with the measure 'Time needed to lie down' from the WQ protocol. The measures included in the WQ protocol concerning 'Comfort around resting' also contribute with more weight in the total score, thereby nullifying the significant agreement in cleanliness.

In the development of the WQ protocol a number of behavioural measures focusing on resting were tested for validity. It was found that both consistency over time and correlation between observers were significant in connection with both 'Time needed to lie down' and 'Collision with housing equipment during lying down' (Brörkens *et al* 2009). Getting-up behaviour could easily be replaced by these measures in the DCF protocol without significantly increasing the time required, since only recording of six lying down events per farm is required (Welfare Quality® 2009).

Principle 3 — Good Health

The two protocols were in agreement over Principle 3. In the DCF protocol four measures are included in P3: lameness, integument alterations, coat condition, and hoof condition. In the WQ protocol 14 measures are included in P3: lameness, integument alterations, coughing, nasal discharge, ocular discharge, hampered respiration, diarrhoea, vulvar discharge, milk somatic cell count, mortality, dystocia, downer cows, disbudding/dehorning, and tail docking (see Table 1).

There were no farms that reached either the warning threshold or the alarm threshold for coughing, nasal discharge, ocular discharge, hampered respiration, diarrhoea or vulvar discharge (Welfare Quality® 2009). One farm reached the warning threshold of 2.75% for downer cows, and four reached the warning threshold of 2.75% for dystocia. With mortality, 15 farms reached the warning threshold of 2.25% and 25 reached the alarm threshold of 4.5%. With milk somatic cell count, eleven farms reached the warning threshold of 8.75% and 29 reached the alarm threshold of 17.5%. The criterion 'Absence of pain induced by management procedures' in the WQ protocol is irrelevant in Danish conditions, because in Denmark the disbudding/dehorning of cattle is only allowed using thermocautery and anaesthesia (Anonymous 1997), and routine tail docking is prohibited (Anonymous 2003). Although only two measures of good health are shared by the protocols, they agree. As the coughs counted in the WQ protocol were found to be few and diffuse they can probably be ascribed to food or water going down the wrong way, rather than illness. Animals housed in sick pens were not included in any of the protocols; this could be one explanation why nasal discharge, ocular discharge, hampered respiration, and vulvar discharge, were not seen. In the WQ protocol, diarrhoea is defined as having loose, watery faeces below the tail head on both sides of the tail; the area affected should be at least the size of a hand. In the study, only nine animals representing six farms were classified as having diarrhoea. Diarrhoea can affect welfare and is an important measure to include (Canali *et al* 2009). However, during development of the WQ protocol, it was recommended that the consistency of faeces be looked at, as opposed to watery faeces below the tail head, as it represents a better indicator of diarrhoea (Canali *et al* 2009). If a measure for diarrhoea is to be included then we suggest that the initial recommendation is adopted.

It can also be argued that it is necessary to include dystocia, mortality, and milk somatic cell count, in the DCF protocol, as these measures are important to animal well-being, and as a rather large percentage of the farms reached the thresholds set by the WQ protocol. In Denmark, databases including these parameters exist; data could thus be easily extracted. The introduction of these measures will therefore not lengthen the time needed to perform the DCF protocol. It can further be argued that it is relevant to replace the lameness scale in the DCF protocol with the one used in the WQ protocol, as this measure has been validated in the latter protocol (Leach *et al* 2009b).

Principle 4 — Appropriate behaviour

The two protocols were also in agreement over Principle 4. One measure (confidence) is used in the DCF protocol, whereas four measures (agonistic behaviour, access to pasture, avoidance distance and Qualitative Behaviour Assessment [see Table 1]) are used in the WQ protocol.

All 44 farms received the highest index score, of 100, for agonistic behaviour (WQ). Only two of the five measures used to address agonistic behaviour in the WQ protocol were recommended as useful (Laister *et al* 2009a). The non-existent variability may suggest that these measures can be excluded under Danish conditions.

Not all farms gave cows access to pasture, but this did not affect the overall agreement. Access to pasture is not an animal-based measure. On pasture, cows can perform aspects of natural behaviour, but high yielding cows may face problems meeting their dietary requirements. Weather conditions and the opportunity to find shelter are important for the cow, and studies have found that cows show a preference for indoor housing in poor weather conditions (Charlton *et al* 2011). Access to pasture is often included in welfare assessment protocols relating to dairy cows, and consumers appear to prefer animal welfare-friendly production systems that allow natural living when asked about welfare (Terragani & Torjusen 2007; Napolitano *et al* 2010). However, there is a risk that clarity and transparency will be lost when a protocol which is set up to obtain an indication of whether the cow itself shows positive or negative reactions to its treatment is mixed up with measures of the resources that have been provided with the aim of improving welfare.

Avoidance distance was measured differently in the two protocols: in the DCF protocol it was measured in the barn (AD), and in the WQ protocol it was measured at the feeding place (ADF). During development of the WQ protocol both AD and ADF were found to be reliable and

https://doi.org/10.7120/09627286.23.1.081 Published online by Cambridge University Press

valid (Windschnurer *et al* 2009). With the WQ protocol, a decision was made to use the ADF. The ADF seems more applicable, and therefore the recommendation here is to use it instead of the AD in the DCF protocol.

Qualitative Behaviour Assessment (QBA) is a relatively new method which uses a qualitative approach to sum up the demeanour and context of the animals being observed (Wemelsfelder 1997; Wemelsfelder *et al* 2000). To date, little is known about this method and its quality. A number of studies have shown good inter-observer reliability but, equally, others have failed to do so (Brscic *et al* 2009; Minero *et al* 2009; Bokkers *et al* 2012). The results of studies investigating correlations between QBA and quantitative measures are also mixed (Napolitano *et al* 2012; Stockman *et al* 2012; Andreasen *et al* 2013).

In the DCF protocol, measures for positive behaviour, such as grooming (social licking) and caudal licking, are included. As these behaviours occurred on all farms they did not affect the result. It is doubtful whether grooming (social licking) and caudal licking should be included in a simplified welfare protocol, since the meaning, or welfare-significance, of these behaviours is, as yet, not greatly understood. A study, performed during the development of the WQ protocol, which examined social licking, supported the idea that social licking does have a calming effect, but this kind of behaviour might also be performed to reduce social tension and individual stress responses, and therefore it cannot be viewed as a solely positive indication (Laister *et al* 2009b).

From the above, we suggest that the original DCF protocol is extended in order to achieve a more all-embracing approach with six new measures: 'Water provision' (P1); 'Time needed to lie down' (P2); 'Collision with housing equipment' (P2); 'Dystocia' (P3); 'Mortality' (P3); and 'Milk somatic cell count' (P3). Further, we recommend that three of the procedures used in the DCF protocol are replaced by the validated procedures used in the WQ protocol; 'BCS' (P1); Lameness scoring' (P3); and 'Confidence' (Avoidance distance) (P4). We also recommended leaving out 'Positive behaviour', ie grooming and caudal licking and 'Getting-up behaviour' from the DCF protocol. These improvements can be done without expanding the time-frame of the DCF protocol.

Below, we present the results reached when correlating the WQ protocol with the extended DCF protocol.

Part 2 — Extending the Danish Cattle Federation welfare protocol

Materials and methods

The study herds and the WQ approach are the same as presented earlier.

The extended DCF protocol now consists of 14 measures and, of these, 13 are animal-based and one is resource-based (see Table 7).

Statistical analysis

The statistical analysis software and the approach used for the WQ protocol were presented earlier.

The extended Danish Cattle Federation Protocol

Six new measures have been included, three measures have changed measuring procedure and two have been excluded.

'Water provision' is incorporated in the protocol. The result reached in the WQ protocol is directly transferred to the extended DCF protocol as the approach used in the WQ protocol has been carefully prepared (Forkman & Keeling 2009).

When looking at 'Time needed to lie down' the raw data retrieved on each farm have been used, this is also the case with 'Collision with housing equipment'. Data concerning 'Dystocia', 'Mortality' and 'Milk somatic cell count' have been retrieved from databases with the same basic considerations used in the WQ protocol (Welfare Quality® 2009).

For each of the three measures ('BCS', 'Lameness' and 'Confidence'), which had the measuring procedure changed to the procedure used in the WQ protocol, the animals inspected in the DCF sample size were sought out in the WQ sample size and the results for these animals were transferred to the extended DCF protocol.

As described earlier, percentiles were found for each of the 14 measures and the farm was ascribed a score from 1–4. Again, 'Lying behaviour' was excluded from the calculation as no animals were found to be lying outside the cubicles.

Comparison of the WQ protocol and the extended DCF protocol

The protocols were compared using Spearman rank correlation. Again, the correlations were performed on the continuous data on the Principles (P1–P4) and on the Overall score (OV).

Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were calculated using a 2×2 table for all four principles (P1–P4) and for the Overall score (OV). For calculation, a farm was considered to have good welfare when classified as either 'Excellent' or 'Enhanced' in the WQ protocol and as either among the 25% best farms or the 25% second-best farms in the extended DCF protocol. A farm was, on the other hand, considered to have poor welfare when classified as either 'Acceptable' or 'Not classified' in the WQ protocol and as either among the 40% second worst farms or the 10% worst farms in the extended DCF protocol.

Results

Using the Spearman rank correlation significant positive correlations were found between all four Principles and also on the Overall score (see Table 3).

OV had a sensitivity of 0.82, a specificity of 0.73, a PPV of 0.75 and an NPV of 0.80. P1 had a sensitivity of 0.71, a specificity of 0.50, a PPV of 0.63 and an NPV of 0.59. P2 had a sensitivity of 0.93, a specificity of 0.47, a PPV of 0.45 and an NPV of 0.93. P3 had a sensitivity of 0.83, a specificity of 0.55, a PPV of 0.23 and an NPV of 0.95. Finally, P4 had a sensitivity of 0.57, a specificity of 0.51, a PPV of 0.18 and an NPV of 0.86.

Discussion

Extending the DCF protocol is a move towards encompassing more aspects of animal welfare and can be done without increasing the time-frame. Significant correlation with the WQ protocol is reached on all tested levels. It is important to consider the validity of measures when replacing existing measures with new ones. The mere fact that there is a good correlation in the current data set does not mean that the two protocols necessarily measure the same thing in all contexts. In the present study, we believe that the measures included in the extended DCF protocol all possess face validity, as is the case with the measures used in the WQ protocol, and they can therefore be argued to measure and reflect welfare. Furthermore, the relatively homogenous population (only Danish dairy herds) secures a uniform context. Despite this argument, only moderate correlation was found looking at P1 and P4. 'Water provision' was in the extended DCF protocol incorporated in C1, which resulted in a moderate correlation in P1. That the correlation is only moderate can, we believe, be ascribed to the fact that the WQ protocol does not take fat animals into consideration, whereas in the extended DCF protocol the score is affected negatively by the occurrence of fat animals. Also, only a moderate correlation was found when looking at P4 (both when comparing WQ to the original and the extended DCF protocol). In the extended DCF protocol, the measure of 'Positive behaviour' was excluded and only confidence (avoidance distance) represented this Principle. Agonistic behaviour (WQ) was, as previously argued, not included, as all farms reached the same score. A possible reason for the moderate correlation for P4 could be that the measures for positive behaviour, ie grooming and caudal licking in the DCF protocol and access to pasture and Qualitative Behaviour Assessment in the WQ protocol, add a perspective not captured by avoidance distance. However, as previously discussed, access to pasture is not an animal-based measure and studies have reached opposing conclusions concerning the animal welfare effect of pasture. Regarding the other measures that were excluded it can be speculated that they give an impression of the cow's mental well-being. However, studies conducted using the measures have, as is the case with access to pasture, given conflicting results (see previous discussion) and, therefore, it is difficult to argue that they should be included in a simple welfare assessment protocol. Research conducted on the measures should, of course, continuously serve to evaluate this and the measures could be included if there turns out to be sufficient evidence for doing so.

In the calculation of sensitivity, specificity, PPV and NPV the results of the WQ protocol are used as a gold standard. When looking at sensitivity, specificity, PPV and NPV the correlations found were supported. High sensitivity was found for OV, P1, P2 and P3 and high NPV was found when looking at OV, P2, P3 and P4. When deciding to use a simplified animal welfare assessment protocol it is important to be aware of the increased risk of both false positives and false negatives, ie the increased risk of scoring farms as having poor welfare when this is not the case and the increased risk of scoring farms as having good welfare when this is not the case. Even though the sensitivity, specificity, PPV and NPV reached for the OV are relatively high the results still indicate the possibility of both false negatives and false positives. However, when looking at the time saved, approximately 6 h on a farm with 200 dairy cows, and the objective of wanting to have animal-based measures used in practice, we believe that the extended DCF protocol correctly balances the demand of practicability against the value as a diagnostic test.

Another potential point of criticism when looking at the DCF protocol is the sample size used. In WQ, 3,051 cows were inspected which corresponds to 38% of the population. In the DCF protocol (extended and original), 1,288 cows were inspected which corresponds to 16% of the population. Increasing the sample size will increase the time requirements for the DCF protocol. Again, with a reference to the similarities in the population and the aim of identifying farms with overall poor welfare, we believe that the power lost when using a smaller sample size is defendable.

However, to further support the above findings a study using the whole extended DCF protocol and the WQ protocol should be conducted.

For summarising the DCF protocol four percentiles were used. The percentiles are based on the population in which the welfare assessment is performed, and therefore the cut-offs found are relative and will change when general welfare levels in the population change. In the WQ protocol, by contrast, the scores assigned to each farm are absolute and independent of the population (Welfare Quality® 2009). This makes the WQ protocol a more appropriate tool for comparing farms across different countries, and for use in labeling. However, when it comes to the problem of how to aggregate measures relating to different principles and criteria, the WQ involves a complex method. To deal with various ethical concerns the WQ protocol uses weights that are based on judgments from a number of experts. The weights reflect the importance the experts assign to each measure (Botreau *et al* 2009). Since, in that way, a number of both factual and ethical assumptions will be buried in one aggregation procedure, the result may end up being opaque for most users, and far from simple (Veissier *et al* 2011). In light of this, it might therefore be better to use a more transparent and slightly less advanced method of summarising measures. If a simple, animal-based welfare assessment protocol, such as the DCF protocol, is implemented, the emphasis should be on making the summarising of the protocol independent of the population, but not dependent on weightings and threshold between acceptable and unacceptable welfare conducted in an opaque way.

Both the WQ protocol and the DCF protocol were applied by the same observer (SNA). High inter-observer reliability has been found for many of the measures used in the WQ protocol (Forkman & Keeling 2009), and the WQ programme confirms that assessors can execute the protocol after receiving training (Anonymous 2013c). Many of the

https://doi.org/10.7120/09627286.23.1.081 Published online by Cambridge University Press

measures used in the DCF protocol resemble the measures in the WQ protocol (see Welfare Quality® 2009 and Appendix 1) and, therefore, inter-observer reliability ought to be satisfactory. However, a study confirming inter-observer reliability among DCF protocol observers would improve the standing of the DCF protocol.

A clear difference between the WQ and the DCF protocols is that the latter was developed specifically for Danish conditions, while the former is a more generalised procedure for use in different countries. Care should therefore be taken in interpreting the DCF results as valid across many differing contexts and environments. In further efforts to develop simple protocols, the range of conditions over which they are to be applied must be taken into consideration.

The DCF protocol was developed by the professional organisation representing cattle farming in Denmark, the Danish Cattle Federation. The DCF represents the interests of Danish cattle farmers, primarily dairy cattle and, as Ingemann *et al* (2009) have pointed out, this could potentially provide "a conflict between the best possible welfare for the animal and the best possible income for the farmers". In view of this, a bias in the DCF's chosen measures might be suspected. The findings in this study do not, however, support this suspicion.

Animal welfare implications and conclusion

This study shows that a simplified, animal-based welfare assessment protocol for dairy cattle in Danish conditions can be constructed. The protocol would be practicable and need not be too time consuming in its application.

Two out of four principles were found to correlate significantly when the original DCF and the more complex WQ protocols were compared. When extending the DCF protocol with several measures making the approach encompass more aspects of animal welfare, correlation between the two protocols was obtained on all levels (Principles and Overall score). The extended DCF protocol can be used without increasing the time needed to carry out the welfare assessment. We therefore conclude that a procedure based on the less time consuming, extended DCF protocol can be used to summarise the more comprehensive WQ protocol in Danish conditions. The adoption of the extended DCF protocol, or simple protocols resembling it, will save both time and money, and this ought to encourage uptake by the dairy industry. More widespread use of animal welfare assessment at farm level could be a way to improve cow welfare. This will not only benefit the cows; it may also improve consumer confidence.

Acknowledgements

The authors wish to thank the farmers who participated in this study. We are also very grateful to Peter Stamp Enemark from the Danish Cattle Federation for his ongoing help and his advice on the DCF protocol.

References

Andreasen SN, 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. *Applied Animal Behaviour Science 143*: 9-17. http://dx.doi.org/10.1016/j.applanim.2012.11.013

Anonymous 1997 *Bekendtgørelse om afhorning af dyr*. https://www.retsinformation.dk/Forms/R0710.aspx?id=85970. [Title translation: Order concerning the dehorning of animals]

Anonymous 2003 *Bekendtgørelse om halekupering og kastration af dyr*. http://www.retsinformation.dk/Forms/R0710.aspx?id=1572. [Title translation: Order concerning tail docking and the castration of animals]

Anonymous 2013a *Fråga Kon*. http://www.svenskmjolk.se/Mjolkgarden/Djurvalfard/Service-och-radgivning/Fraga-kon/. [Title translation: Ask the cow]

Anonymous 2013b *Danish Metereological Institute*. http://www.dmi.dk/dmi/saadan_er_sommeren

Anonymous 2013c *The Welfare Quality® homepage*. http://www.welfarequality.net/everyone/43299/7/0/22

Blokhuis H, Miele M, Veissier I and Jones B 2013 *Improving farm animal welfare. Science and society working together: the Welfare Quality approach*. Wageningen Academic Publishers: The Netherlands. http://dx.doi.org/10.3920/978-90-8686-770-7

Blokhuis HJ, Veissier I, Miele M and Jones B 2010 The Welfare Quality® project and beyond: safeguarding farm animal well-being. *Acta Agriculturae Scandinavica, Section A - Animal Science 60*: 129-140

Bokkers EAM, de Vries M, Antonissen ICMA and de Boer IJM 2012 Inter- and intra-observer reliability of experienced and inexperienced observers for the Qualitative Behaviour Assessment in dairy cattle. *Animal Welfare 21*: 307-318. http://dx.doi.org/10.7120/09627286.21.3.307

Botreau R, Veissier I, Butterworth A, Bracke MBM and Keeling LJ 2007 Definition of criteria for overall assessment of animal welfare. *Animal Welfare 16*: 225-228

Botreau R, Veissier I and Perny P 2009 Overall assessment of animal welfare: strategy adopted in Welfare Quality® . *Animal Welfare 18*: 363-370

Brörkens N, Plesch G, Laister S, Zucca D, Winckler C, Minero M and Knierim U 2009 Reliability testing concerning behaviour around resting in cattle in dairy cows and beef bulls. *Assessment of Animal Welfare Measures for Dairy Cattle, Beef Bulls and Veal Calves. Welfare Quality Reports No 11* pp 7-24. Sixth Framework Programme: Cardiff University, UK

Brscic M, Wemelsfelder F, Tessitore E, Gottardol F, Cozzi G and Van Reenen CG 2009 Welfare assessment: correlations and integration between a Qualitative Behavioural Assessment and a clinical/health protocol applied in veal calves farms. *Italian Journal of Animal Science 8*: 601-603

Canali E, Whay HR and Leach KA 2009 Cattle health status. *Assessment of Animal Welfare Measures for Dairy Cattle, Beef Bulls and Veal Calves. Welfare Quality Reports No 11* pp 77-88. Sixth Framework Programme: Cardiff University, UK

Charlton GL, Rutter SM, East M and Sinclair LA 2011 Preference of dairy cows: indoor cubicle housing with access to a total mixed ration vs access to pasture. *Applied Animal Behaviour Science 130*: 1-9. http://dx.doi.org/10.1016/j.applanim.2010.11.018

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

de Vries M, Engel B, den Uijl I, van Schaik G, Dijkstra T, de Boer IJM and Bokkers EAM 2013 Assessment time of the Welfare Quality® protocol for dairy cattle. *Animal Welfare 22*: 85-93. http://dx.doi.org/10.7120/09627286.22.1.085

Farm Animal Welfare Committee 2013 *Five Freedoms*. http://www.fawc.org.uk/freedoms.htm

Forkman B and Keeling L 2009 *Assessment of Animal Welfare Measures for Dairy Cattle, Beef Bulls and Veal Calves. Welfare Quality Reports No 11*. Sixth Framework Programme: Cardiff University, UK

Hemsworth PH, Barnett JL, Beveridge L and Matthews LR 1995 The welfare of extensively managed dairy cattle: a review. *Applied Animal Behaviour Science 42*: 161-182. http://dx.doi.org/10.1016/0168-1591(94)00538-P

Igbokwe IO 1997 The effects of water deprivation in livestock ruminants: an overview. *Nutrition Abstracts and Reviews 50*: 905-914

Ingemann R, Sandøe P, Enemark P and Forkman B 2009 Conflicting goals of welfare assessment schemes: a case study. *Animal Welfare 18*: 487-495

Johnsen PF, Johannesson T 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. http://dx.doi.org/10.1080/09064700131692302 7

Laister S, Brörkens N, Lolli S, Zucca D, Knierim U, Minero M, Canali E and Winckler C 2009a Reliablity of measures of agonistic behaivour in dairy and beef cattle. *Assessment of Animal Welfare Measures for Dairy Cattle, Beef Bulls and Veal Calves, Welfare Quality Reports No 11* pp 95-112. Sixth Framework Programme: Cardiff University, UK

Laister S, Regner A-M, Zenger K, Winckler C, Brörkens N, Quast R and Knierim U 2009b Validation of social licking as an indicator for positive emotions. *Assessment of Animal Welfare Measures for Dairy Cattle, Beef Bulls and Veal Calves, Welfare Quality Reports No 11* pp 113-123. Sixth Framework Programme: Cardiff University, UK

Leach KA, Knierim U and Whay HR 2009a Condition scoring for dairy and beef cattle and veal calves. *Assessment of Animal Welfare Measures for Dairy Cattle, Beef Bulls and Veal Calves, Welfare Quality Reports No 11* pp 1-6. Sixth Framework Programme: Cardiff University, UK

Leach KA, Winckler C and Whay HR 2009b Lameness in dairy and beef cattle and veal calves. *Assessment of Animal Welfare Measures for Dairy Cattle, Beef Bulls and Veal Calves, Welfare Quality Reports No 11* pp 35-41. Sixth Framework Programme: Cardiff University, UK

Miele M, Veissier I, Evans A and Botreau R 2011 Animal welfare: establishing a dialogue between science and society. *Animal Welfare 20*: 103-117

Minero M, Tosi MV, Canali E and Wemelsfelder F 2009 Quantitative and qualitative assessment of the response of foals to the presence of an unfamiliar human. *Applied Animal Behaviour Science 116*: 74-81. http://dx.doi.org/10.1016/j.applanim.2008.07.001

Napolitano F, De Rosa G, Grasso F and Wemelsfelder F 2012 Qualitative behaviour assessment of dairy buffaloes (*Bubalus bubalis*). *Applied Animal Behaviour Science 141*: 91-100. http://dx.doi.org/10.1016/j.applanim.2012.08.002

Napolitano F, Girolami A and Braghieri A 2010 Consumer liking and willingness to pay for high welfare animal-based products. *Trends in Food Science & Technology 21*: 537-543. http://dx.doi.org/10.1016/j.tifs.2010.07.012

Roche JR, Friggens NC, Kay JK, Fisher MW, Stafford KJ and Berry DP 2009 Invited review: body condition score and its association with dairy cow productivity, health and welfare. *Journal of Dairy Science 92*: 5769-5801. http://dx.doi.org/10.3168/jds.2009-2431

Roenfeldt S 1998 You can't afford to ignore heat stress. *Dairy Herd Management 35(5)*: 6-12

Rushen J and de Passillé AMB 1992 The scientific assessment of the impact of housing on animal welfare: a critical review. *Canadian journal of animal science 72*: 721-743. http://dx.doi.org/10.4141/cjas92-085

Stockman CA, McGilchrist P, Collins T, Barnes AL, Miller D, Wickham SL, Greenwood PL, Cafe LM, Blache D, Wemelsfelder F and Fleming PA 2012 Qualitative Behavioural Assessment of Angus steers during pre-slaughter handling and relationship with temperament and physiological responses. *Applied Animal Behaviour Science 142*: 125-133. http://dx.doi.org/10.1016/j.applanim.2012.10.016

Terragani L and Torjusen H 2007 Norway. *Consumers' Views About Farm Animal Welfare, Part 1: National Reports Based on Focus Group Research. Welfare Quality Reports No 4* pp 253-322. Sixth Framework Programme: Cardiff University, UK

Veissier I, Jensen KK, Botreau R and Sandøe P 2011 Highlighting ethical decisions underlying the scoring of animal welfare in the Welfare Quality® scheme. *Animal Welfare 20*: 89-101

Waiblinger S, Knierim U and Winckler C 2001 The development of an epidemiologically based on-farm welfare assessment system for use with dairy cows. *Acta Agriculturae Scandinavica Section A - Animal Science 30*: 73-77. http://dx.doi.org/10.1080/09064700131692310 8

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

Wemelsfelder F 1997 The scientific validity of subjective concepts in models of animal welfare. *Applied Animal Behaviour Science 53*: 75-88. http://dx.doi.org/10.1016/S0168-1591(96)01152-5

Wemelsfelder F, Hunter EA, Mendl MT and Lawrence AB 2000 The spontaneous qualitative assessment of behavioural expressions in pigs: first explorations of a novel methodology for integrative animal welfare measurement. *Applied Animal Behaviour Science 67*: 193-215. http://dx.doi.org/10.1016/S0168-1591(99)00093-3

Windschnurer I, Schmied C, Boivin X and Waiblinger S 2009 Assessment of human-animal relationships in dairy cows. *Assessment of Animal Welfare Measures for Dairy Cattle, Beef Bulls and Veal Calves, Welfare Quality Reports No 11* pp 137-152. Sixth Framework Programme: Cardiff University, UK

https://doi.org/10.7120/09627286.23.1.081 Published online by Cambridge University Press