© 2011 Universities Federation for Animal Welfare
The Old School, Brewhouse Hill, Wheathampstead,
Hertfordshire AL4 8AN, UK

Animal Welfare 2011, 20: 239-251
ISSN 0962-7286

Assessment of sheep welfare using on-farm registrations and performance data

SM Stubsjøen*[†‡], L Hektoen[‡], PS Valle[†§], AM Janczak[†] and AJ Zanella[†#¶]

[†] Norwegian School of Veterinary Science, Department of Production Animal Clinical Sciences, PO Box 8146 Dep, NO-0033 Oslo, Norway
[‡] Animalia, Norwegian Meat and Poultry Research Centre, PO Box 396, Økern, NO-0513 Oslo, Norway
[§] Molde University College, PO Box 2110, NO-6402 Molde, Norway
[#] The Norwegian University of Life Sciences, Department of Animal and Aquacultural Sciences, PO Box 5003, NO-1432 Aas, Norway
[¶] Animal Behaviour and Welfare, Sustainable Livestock Systems Group, Scottish Agricultural College (SAC), SAC Roslin Building, Easter Bush, Midlothian EH25 9P6, UK
* Contact for correspondence and requests for reprints: SolveigMarie.Stubsjoen@nvh.no

Abstract

*Farm animal welfare is a societal concern, and the need exists for scientific protocols to assess welfare. This paper describes the development of a protocol to assess the welfare of sheep (*Ovis aries*) and its application in 36 farms in Norway. There were two parts to the protocol; the animal- and resource-based measurements obtained during farm visits, and the analysis of production data. Data collection took place during visits to 36 farms in the lambing season (April-May) in 2007 (n = 11) and 2008 (n = 25). A fear test was conducted, and ewes were scored on a scale from 0 to 3. Forty-one percent of the ewes tested had a fear score of 3, indicating the lowest level of fear. Mean (± SD) fear score across farms were 1.9 (± 0.5). Higher fearfulness was found to be associated with lower ewe body condition scores (BCS). Mean (± SD) BCS across farms was 2.6 (± 0.6). A relatively large proportion of the ewes had a BCS of 2 (41%), which may be associated with an increased risk of nutritional stress, disease and low productivity. Eight farms had more than 5% (range 5.4–24.4%) of lamb carcases categorised in the lowest conformation class, which may be an indication of a welfare problem. This study is the first step in the development and validation of a welfare assessment protocol for sheep, and further research is needed to assess the overall reliability of the protocol.*

Keywords: *animal-based measures, animal welfare, production, resource-based measures, sheep, welfare assessment protocol*

Introduction

Sheep production is of major economic importance in many countries and has been the subject of less industrialisation than many other forms of livestock production. Reduced economic output may, however, be a risk factor for sub-optimal health, handling and poor welfare situations, since there is little room for input resources per animal. Public concern about farm animal welfare has steadily increased during recent years. The majority of participants in population surveys carried out in seven European countries (2005) believed that farm animals feel pain like humans, indicating an acknowledgement of farm animals as live, sentient beings (Kjaernes & Lavik 2007). Increasing demand from customers for humane production has put pressure on livestock industries to improve and provide evidence of the welfare status of their animals. Therefore, there is a need for scientifically based welfare assessment protocols.

Sheep undergo painful husbandry procedures in many countries, such as castration and tail-docking of lambs (Molony & Kent 1997). This species also experiences a wide range of diseases and tissue injuries, including mastitis, footrot and fly-strike. Sheep are stoic creatures, and they do not display overt behavioural signs of distress and pain. Human observers may also lack the ability or skills to identify behaviours indicative of sub-optimal welfare in sheep.

Examples of existing on-farm monitoring systems include the Tiergerechtheitsindex (TGI) developed in Austria, the Bristol Welfare Assurance Programme (BWAP 2009) and the Welfare Quality® (2009) project protocols. The TGI system focuses mainly on resource-based measures (eg floor type and space allocation). Today, there is considerable agreement to use mainly animal-based measures when assessing animal welfare (Keeling & Veissier 2005). The Welfare Quality® welfare assessment protocols and the Bristol welfare assurance programme (BWAP) protocols are developed for the assessment of cattle, poultry and pig welfare. These protocols focus essentially on animal-based measures. Many of the welfare measures applied in the

Welfare Quality® and BWAP protocols can be regarded as sufficiently valid, reliable and feasible. However, the overall reliability of these existing protocols may be further improved by refinement of definitions, recording methods and training. All welfare measures incorporated into the Welfare Quality® protocol possess face validity, but for most of them construct or criterion validity have not been demonstrated (Knierim & Winckler 2009). Also, time constraints are a major concern. Currently, 6 h is required to assess a herd of 60 dairy cows using the Welfare Quality® protocol (Knierim & Winckler 2009), which decreases the feasibility of using the protocol.

Although increased attention is paid to development of on-farm welfare monitoring protocols (eg http://www.wafl2008.be/site/index.php), there is no existing protocol for on-farm assessment of welfare in sheep. The need for a comprehensive and uniform method for inspectors to assess welfare is no less for this species than for other farm animals. The Royal Society for the Prevention of Cruelty to Animals (RSPCA) and Certified Humane have standards for sheep which can be considered as minimum requirements. Protocols cannot automatically be transferred from one species to another. For instance, the behaviour of sheep is quite different from the behaviour of cattle or pigs. Hence, specific protocols to assess sheep welfare need to be developed.

The essential needs of domesticated animals include physical needs for food, water, thermal comfort and rest, the possibility to express preferred behaviours, such as foraging and exploration, as well as freedom from pain, fear and distress. These requirements often remain unfulfilled in modern farming practice, thereby leading to many welfare, production and product quality problems (Blokhuis *et al* 2003). Improvements in animal welfare may be achieved through a process based on the assessment of animal welfare using welfare assessment systems, identification of risk factors potentially leading to welfare problems, and interventions in response to the risk factors (Whay 2007).

Three different views on animal welfare exist and must be taken into account when developing an animal welfare assessment protocol (Broom 1991). In general, these different views are focused on three broad questions: i) is the animal functioning well; ii) is the animal feeling well; and iii) is the animal able to live a reasonably natural life (Duncan & Fraser 1997). Also, a range of indicators must be evaluated and combined in order to assess welfare (Broom 1991). Animal-based observations are a direct assessment of animals' health and behaviour, and provide the most direct insight into how animals are coping with their environment. Resource-based observations focus on what has been provided for the animals, such as shelter, comfort, space allowance and nutrition, and is a more indirect measure of animal welfare (Whay 2007). Welfare Quality® identified four principles and 12 areas of concern that should be adequately covered in welfare assessment systems, corresponding to questions of whether the animals are properly fed and supplied with water, properly housed, healthy and exhibit behaviours that reflect optimised emotional states (Blokhuis 2008). While emphasising the Five Freedoms (Farm Animal Welfare Council 1993) and the four principles of Welfare Quality®, we included animal- and resource-based parameters in the welfare assessment protocol.

One of the key recommendations of the Brambell Committee in 1965 was that intensively housed livestock should be free from fear. Fear is a potent stressor that elicits activation of the hypothalamic-pituitary-adreno-cortical axis and suppresses the immune system, growth, productivity and feed conversion (Jones 1997). High levels of fearfulness may have negative consequences both for the welfare of the animals and the income of the producer, and is a serious threat to welfare (Jones 1997; Rushen *et al* 1999; Waiblinger *et al* 2006). The development of a positive human-animal relationship (low levels of fear or high levels of confidence in people) may be beneficial. The presence of a familiar human may calm the animals in potentially aversive situations, thus reducing distress and the risk of injury to the animal and the stockperson (Waiblinger *et al* 2006).

We developed a welfare assessment protocol for sheep based on the Five Freedoms, established welfare assessment protocols for dairy cattle (Welfare Quality® and Bristol Welfare Assurance Programme) and welfare standards for sheep outlined by RSPCA and Certified Humane. Secondary data from databases of performance were included as additional outcome measures relevant to animal welfare assessment.

The aim of this study was to contribute to the development of an on-farm welfare assessment protocol for sheep and to assess its application in 36 Norwegian sheep flocks. There are two parts to the protocol; the observations and measurements made by two observers during the farm visits and the analysis of records of production data.

Materials and methods

A protocol for on-farm assessment of sheep welfare was developed based on the Five Freedoms, established welfare assessment protocols for dairy cattle (Welfare Quality® and Bristol Welfare Assurance Programme) and welfare standards for sheep outlined by RSPCA and Certified Humane. The observations and records included in the protocol were selected after consultations with an expert group, on the basis that the parameters should be important for sheep welfare. Animal- and resource-based measurements, as well as data based on production records, were included in the protocol and are listed in Table 1.

Description of the farms

Thirty-six farms distributed over three different geographical regions (north, south-west and the mountain region in the mid-east of Norway) were observed during one visit at each farm in the lambing season (April-May) in 2007 (n = 11) and 2008 (n = 25). The lambing season was selected due to the increased density of animals in this period enhancing the risk of contagious diseases and other welfare-related problems. Farms were recruited through

Table 1 List of the animal- and resource-based measurements and production data included in the welfare assessment protocol. The list was created based on the Five Freedoms and established welfare assessment protocols (Welfare Quality®, Bristol Welfare Assurance Programme [BWAP], Royal Society for the Prevention of Cruelty to Animals' welfare standards for sheep, and Certified Humane's animal care standards for sheep).

Animal-based measurements	Resource-based measurements	Data based on production records
Body condition score	Size of pens	Slaughter weight
Animal appears sick/dull	Number and size of animals in pens	Carcase classification
Lameness	Size of trough space	Fat class
Cleanliness	Temperature	
Diarrhoea	Surface temperature of lying area	
Skin lesions	Relative humidity	
Skin irritation	Lighting	
Swollen joints	Draught	
Coughing	Ammonia and CO_2	
Eye abnormalities	Solid lying area for lambs	
Nasal discharge	Sharp edges or protrusions	
Udder (inflammation)	Hygiene lying area	
Callus on carpus	Hygiene through space	
Ear-tag (in place or torn out)	Water (access and hygiene)	
Fear	Food (access and subjective assessment of quality	
Human-animal relationship		

random sampling from lists of producers obtained from abattoirs in these three regions. Farmers were contacted by telephone and asked whether they wanted to participate in the study. Eleven farmers were then included in the study from the region of Nord-Østerdal (mid-east; studied in 2007), 15 from Rogaland (south west; studied in 2008) and 10 from Sortland (north; studied in 2008). Data were collected over a two-week period in April and May, 2007 and during a five-week period in April and May, 2008. On the selected farms, the number of ewes varied from 32 to 412. Thirty farms had the Norwegian White breed, while six farms had a combination of Norwegian White, Spælsau or Texel. The age of the ewes ranged from one to eight years.

Animal- and resource-based measures

During the farm visits, the flock was observed to detect signs of clinical disease, lameness and coughing. In our study, sheep were not reared in one group per farm, but were kept in different pens. The number of pens varied according to the size of the farm. Due to climatic conditions in Norway, sheep are typically housed in pens with slatted flooring and a space allowance of 0.7–0.9 m^2 per animal during the winter months (Bøe & Simensen 2003). Ten animals were randomly selected at each farm, and a clinical examination was performed by a veterinary surgeon. In general, one or two animals were examined in each pen, but this varied according to flock size. All the 360 ewes inspected were included in the analyses. Resource-based measurements, such as relative humidity, draught and temperature, were measured several times (range: 3–27 times) during the visits in the pens where animals were kept. The number of measurements at each farm varied due to the different farm sizes. Indicators, such as ammonia and CO_2 concentration and lighting were measured. Each variable was measured in different pens where animals were kept in the farm, and a mean value was obtained. Since the welfare assessments were carried out during the lambing season, when the number of animals increases rapidly, the animals were typically housed in a variety of different places at the farm. The behaviour of the ewes was observed as described below in order to assess the human-animal relationship and level of fear of humans.

There are no animal-based indicators of positive emotions included in the protocol, mainly due to the lack of feasible indicators (Boissy *et al* 2007). Also, resting time, aggressive social interactions and the synchrony of resting behaviour in ewes were not measured, mainly as a result of time constraints. All the observations were performed under indoor conditions. The time required to perform the welfare assessment was between 3 and 5 h. Two observers conducted the measurements. Both were veterinary surgeons with clinical experience from veterinary practice. The fear testing was performed by an ethologist.

Fear of humans

The ewes' response to an unfamiliar person was assessed with the assumption that their responses to the unfamiliar person would reflect, to a certain extent, the way in which

https://doi.org/10.1017/S0962728600002724 Published online by Cambridge University Press

Table 2 Point scale (5 and 15) used for EUROP conformation and fat class.

Confirmation class, 5		15	15
E: Excellent	All profiles convex to extremely convex; exceptional muscle development	E+	15
		E	14
		E–	13
U: Very good	Profiles on the whole convex; very good muscle development	U+	12
		U	11
		U–	10
R: Good	Profiles on the whole straight; good muscle development	R+	9
		R	8
		R–	7
O: Fair	Profiles straight to concave; average muscle development	O+	6
		O	5
		O–	4
P: Poor	Profiles concave to very concave; poor muscle development	P+	3
		P	2
		P–	1
Fat class			
5: Very high	Carcase thickly covered with fat; heavy fat deposits in the thoracic cavity	5+	15
		5	14
		5–	13
4: High	Flesh covered with fat, but on the hindquarter and shoulder still partly visible; some distinctive fat deposits in the thoracic cavity	4+	12
		4	11
		4–	10
3: Average	Flesh, with exception of the hindquarter and shoulder, almost everywhere covered with fat; slight deposits of fat in the thoracic cavity	3+	9
		3	8
		3–	7
2: Slight	Slight fat cover, flesh visible almost everywhere	2+	6
		2	5
		2–	4
1: Low	None up to low fat cover	1+	3
		1	2
		1–	1

they would respond to the stockperson. Fear testing of ewes was performed at 25 farms in 2008, and took place approximately two hours after morning feeding as suggested by Lankin (1997) and Erhard *et al* (2004). A minimum of 20 ewes from each farm were tested. The standardised fear test was conducted by the same test person at all farms. The methodology used for fear testing was a modification of methods validated by Lankin (1997). The test person walked along the feeding area in front of the pens when the ewes were standing, giving them concentrated feed in the feeding area (feed trough), at the front of the pens from a bucket that was borrowed from the farmer. The bucket was then put on the floor, and the test person went calmly in the opposite direction. While moving past the animals, the test person attempted to mark the ewes on the head, using finger-paint on a sponge attached to the end of a broom handle. The bucket was then collected and the test person returned to the starting point, provided the ewes with more concentrated feed and tried to mark them on the back. Finally, this procedure was performed a third time while trying to mark the ewes on the rump. Each ewe received a maximum of one mark on each body part and ewes that avoided the human received marks on only the body part that could be reached. After completing the test, the test person recorded the ear-tag number and number of markings on each individual ewe in each pen. The number

https://doi.org/10.1017/S0962728600002724 Published online by Cambridge University Press

of markings on each ewe varied from zero to three. Ewes with no marks had showed the most pronounced avoidance reactions, and represented the most fearful sheep in the flock. Ewes with three marks were able to be touched on both the head, back and rump, and were thus the least fearful individuals.

Farmer-animal relationship

To test the farmer-animal relationship at 36 farms, the farmer was asked to enter different pens and mark randomly selected ewes in each pen using a marker pen. Marking is a common procedure in Norwegian farms. Ten ewes were tested at each farm. The ewes' response to the farmer was categorised into four groups: 3) behaved calmly when approached; 2) some avoidance; 1) marked avoidance and struggling to escape; and 0) attempts to escape by jumping out of the pen. An average score from each farm was obtained. There were one or two stockpeople at each farm, and the animals could therefore have different relationships with stockpersons. The person who interacted most commonly with the animals was asked to mark them, in order to minimise this confounding effect.

Secondary data of performance

Secondary recordings from databases of performance at Animalia, Norwegian Meat and Poultry Research Centre were analysed and included as additional welfare indicators. These databases contain individual information on slaughter weight, fat class and classification of the carcase. Carcase classification of sheep in Norway, as in the European Union, is based on the EUROP carcase classification system (Council Regulation [EEC] No 2137/92 1992; Commission Regulation [EEC] No 461/93 1993). Conformation class describes carcase shape in terms of convex or concave profiles and is intended to indicate the amount of flesh (meat) in relation to bone. Fat class describes the amount of visible subcutaneous fat on the outside of the carcase (Fisher & Heal 2001). Carcases are classed from 1 to 15 (Table 2).

We compared assessments from 16 farms; 8 farms with the best mean conformation class and 8 farms which had more than 5% of lamb carcases categorised as conformation class P (1–3) in order to qualitatively investigate if there were patterns in the measurements which could explain the occurrence of thin or well-conformed lambs at slaughter. A quantitative analysis was also performed.

Inter-observer agreement

In order to develop a welfare protocol, it is important that the measured parameters have a high reliability. Therefore, a model for evaluation of the repeatability of parameters used by the two assessors was developed. After theoretical and practical training, the assessors examined two groups of ten animals independently. The results were used to identify the repeatability of the animal-based measurements. We used weighted kappa statistics to assess the inter-rater observer agreement, defined as the level of agreement beyond what would have been expected by chance (Dohoo *et al* 2003).

Statistical analysis

Descriptive analysis of the data was performed in Excel® before data were transferred to STATA® (SE/10 for Windows, StataCorp, College Station, TX, USA). A Pearson correlation analysis was performed to assess co-linearity between all factors. Based on this analysis, three measures were selected as outcomes; one production measure (body condition score) and two measures of affective state (fear of humans and farmer-animal relationship). Possible associations between the three outcomes and selected animal- and resource-based measures (listed in Table 1) were investigated. Multiple regression analyses were performed, with one linear (fear of humans) and two logistic (body condition score and farmer-animal relationship) regression models built in STATA®. In all regression models, linearity was investigated with basic graphs, and each explanatory variable was explored against each outcome individually. All models were clustered on herd level to account for any herd effect. In the logistic regression models, body condition score (BCS) was classified as 'poor body condition score' (0: BCS 1 and 2) or 'good body condition score' (1: BCS 3 and 4) and the farmer-animal relationship was categorised as 'some avoidance' and 'marked avoidance and struggling to escape' (0) or 'behaved calmly when approached' (1). In the linear regression model, the association between the continuous variable mean fear of human scores and the number of lambs slaughtered from individual farms (indicative of flock size) was first explored using a plot. A non-linear relationship was found between these two variables. Smaller farms have been found to be associated with more positive behaviour towards calves (Lensink *et al* 2000), and we aimed to investigate if fear of humans was related to different farm sizes. The variable, 'number of lambs slaughtered from individual farms' (n) was therefore categorised into three groups representing 17 small ($n < 200$), four medium ($n = 200$–400) and four large ($n > 400$) farms. Kruskal-Wallis equality of populations rank test is a non-parametric method for testing equality of population medians among groups. This test was used to investigate a possible statistical relationship between farm size and fear of humans.

A selection process was carried out on each explanatory variable (listed in Table 1), and variables with $P < 0.2$ were included as candidate variables for the final model. When building the final model a forward selection procedure was used, starting with the variables with lowest P-values from the selection procedure and using the likelihood ratio test to select variables in the final model. Any distortion and confounding could then be observed as each variable was included separately. Confounding and collinear variables were additionally tested in the regression model.

Results

The selected farms were compared to members of the Norwegian Sheep Recording System (NSRS) registered at Animalia, Norwegian Meat and Poultry Research Centre.

Table 3 Mean score of selected animal†- and resource-based measurements and production data in eight farms which had > 5% of lamb carcases categorised as conformation class (CC) P (1–3). Grade P lamb carcases have concave muscle profiles and a low degree of meat.

	Mean score farmer 1–8									
Variable	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**Mean value 36 farms**	**Recommended values**
BCS[a]	2.7	2.9	2.2	2.7	2.5	2.9	2.4	2.6	2.6	3[#]
Cleanliness[b]	1.9	2.2	1.4	1.9	2.5	2.2	1.3	1.3	1.6	1
Skin lesions[c]	1	1	1	1	1	1.2	1.1	1.1	1	1
Skin irritations[d]	1.2	1.2	1	1.2	1	1.1	1.6	1.4	1.2	1
Callus on carpus[e]	1.5	1.8	1.8	1.5	1.4	1.4	1.2	1.3	1.5	1
Lighting (lux)	146.5	175.7	321.9	146.5	352.5	425	192	28.8	264	≥ 100 lux[§]
Relative humidity	75.3	75.6	57.3	75.3	50.7	65.4	67.7	74.4	66	≤ 70%[§]
Surface temperature (°C)	10	7.4	8.9	10	12.5	10.4	12.7	8.8	11	
Draught (m s^{-1})	0.08	0.08	0.15	0.08	0.16	0.18		0.29	0.2	
Ammonia (ppm)	4	4.5	7.9	4	2.7	8	6.5	1.5	5.8	< 10 ppm[§]
CO_2 (ppm)	600	900	1,250	600	500	205	76.7	295	400.2	< 2,500 ppm[§]
Temperature (°C)	11.6	9.45	11.3	11.6	12	10.5	12.7	9.7	11.6	5–25°C
Fear score[‡] (n = 25)	1.2	2	1.1	1.8	0.9	2.9		2.4	1.9	3
Human-animal relationship*	2	3	2	1	2	2	2	2	1.7	1
CC 1–3 (%)	5.4	5.6	7.8	7.9	8.5	9.3	12.4	24.4	2.9	
Fat class 1–3 (%)	3.9	7	13.7	7.4	5.8	6.2	11.1	17.1	3.9	
Slaughter weight (kg)	17.6	22.2	23.9	20	21.7	17.7	20.1	17	21.4	
Number of lambs slaughtered	259	71	51	203	398	97	81	82	186	

† Categories of selected animal- based measurements: [a] Body condition score (BCS), 1: very thin; 2: thin; 3: average; 4: fat; 5: very fat.
[b] Cleanliness, 1: clean; 2: some dirty parts; 3: dirty; 4: very dirty. [c] Skin lesions, 1: no skin lesions; 2: lesions >1 × 1 cm; 3: ulcerations.
[d] Skin irritations, 1: normal; 2: loss of wool; 3: redness/swelling; 4: parasites or flies. [e] Callus on carpus, 1: no callus; 2: callus; 3: callus with ulcerated skin.
‡ Zero indicates the highest level of fear whereas 3 indicates the lowest level.
* Categories of farmer-animal relationship assessment: 3) Behaved calmly when approached; 2) Some avoidance; 1) Marked avoidance and struggling to escape; and 0) Attempts to escape by jumping out of the pen.
# Recommended body condition score (BCS) at lambing (Stubbings 2007).
§ Sevi *et al* 2007.

The mean flock size (130 ewes) was somewhat higher than the national average (70 ewes) for the thirty-six selected flocks. The mean slaughter weight of lamb carcases of the Norwegian White breed recorded in NSRS (n = 255,227) was 20 kg (Anonymous 2008) while the mean slaughter weight of lamb carcases from farms participating in our study was 21.4 kg. The selected farms were therefore found to be representative of Norwegian conditions according to production data.

The study population was relatively homogenous for several variables, and the variables with some degree of variation are therefore presented in Table 3. The prevalence of dull and depressed animals (0.28%), disease (4.4%), lameness (1.4%), diarrhoea (1.7%), swollen joints (1.7%), coughing (0%), eye abnormalities (1.1%), nasal discharge (0.6%) and inflammation of the udder (1.1%) were low. Mean (± SD) BCS across farms was 2.6 (± 0.6). The majority of the ewes had a BCS of 3 (54.7%). However, a relatively large proportion of the ewes (40.6%) had a BCS of 2. Approximately half of the examined ewes were categorised as clean (50.4%). Fifty-four percent of the ewes had calluses on the carpus. There were few skin lesions (3.9%). Eight percent of the ewes had their ear-tag torn out.

An overview of the 36 selected farms is given in Table 4, which shows the mean of selected animal- and resource-based measurements and production measures. There was a large variation in the mean values measured across farms, especially for the resource-based measures. A large variation was found both between and within farms with regards to the availability of solid resting areas for lambs, which is required by Norwegian legislation.

https://doi.org/10.1017/S0962728600002724 Published online by Cambridge University Press

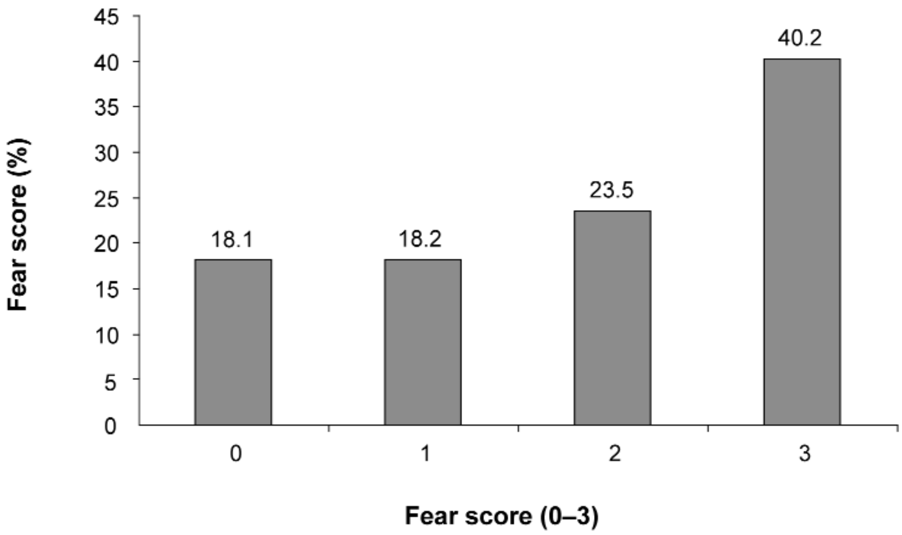
Table 4 Mean (± SD) of selected animal*- and resource-based measurements across 36 farms in three different regions of Norway.

Variable	Observation (n)	Mean (± SD)	Minimum	Maximum	Recommended values
BCS[a]	36	2.6 (± 0.3)	1.9	3.3	3[‡]
Cleanliness[b]	36	1.6 (± 0.4)	1	2.5	1
Skin lesions[c]	36	1.1 (± 0.1)	1	1.5	1
Skin irritation[d]	36	1.2 (± 0.2)	1	1.6	1
Callus on carpus[e]	36	1.5 (± 0.3)	1	2	1
Lux	36	264 (± 289.7)	28.8	1,287.4	≥ 100 lux[§]
Relative humidity (%)	36	66 (± 12.1)	35.3	89	≤ 70%[§]
Surface temperature of lying area (°C)	36	11 (± 2.6)	6	15.2	
Draught ($m\ s^{-1}$)	31	0.2 (± 0.2)	0	0.8	
Ammonia (ppm)	36	5.8 (± 3.4)	1.5	16.5	< 10 ppm[§]
CO_2 (ppm)	36	400.2 (± 433.9)	34	1,600	< 2,500 ppm[§]
Air temperature (°C)	36	11.6 (± 2.9)	5.7	18	5–25°C[§]
Slaughter weight (kg)	36	21.4 (± 1.8)	17	24.4	
Carcase classification[#]	36	7.9 (± 1.3)	4.8	11.1	
Fat class[#]	36	6.2 (± 0.6)	4.4	7.5	

* Categories of selected animal-based measurements: [a] Body condition score (BCS), 1: very thin; 2: thin; 3: average; 4: fat; 5: very fat. [b] Cleanliness, 1: clean; 2: some dirty parts; 3: dirty; 4: very dirty. [c] Skin lesions, 1: no skin lesions; 2: lesions >1 × 1 cm; 3: ulcerations. [d] Skin irritations, 1: normal; 2: loss of wool; 3: redness/swelling; 4: parasites or flies. [e] Callus on carpus, 1: no callus; 2: callus; 3: callus with ulcerated skin. [#] EUROP scale$^{1-15}$. [‡] Recommended body condition score (BCS) at lambing (Stubbings 2007). [§] Sevi *et al* (2007).

Figure 1

Percentage of ewes (n = 912) scored in different fear categories at 25 Norwegian sheep farms. Zero indicates the highest level of fear whereas 3 indicate the lowest level.



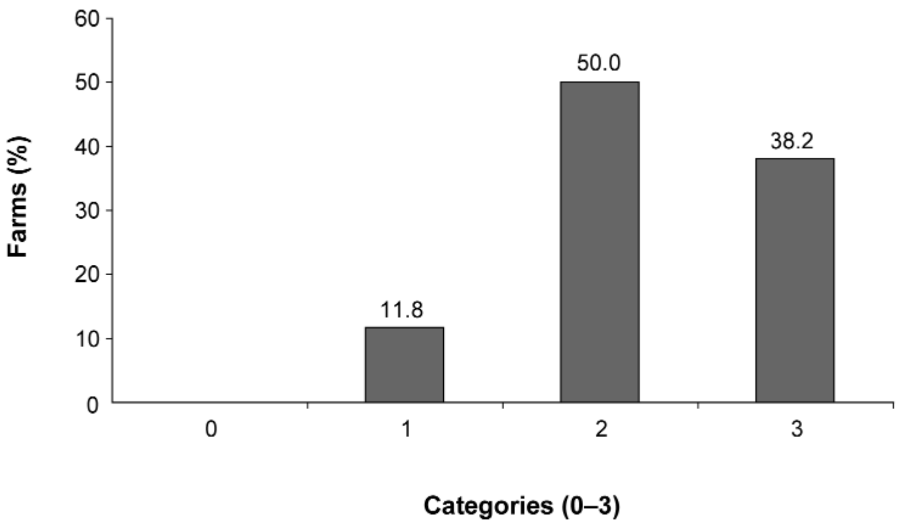
Forty-one percent of the ewes tested at 25 farms had a fear score of 3, which indicates the lowest level of fear (Figure 1). Mean (± SD) fear score across farms were 1.9 (± 0.5), with a range from 0.9 to 2.9.

A farmer-animal relationship score of 2 was obtained in 50% of the farms, which indicates that the ewes showed some avoidance when being marked. However, 38.2% of the farms obtained a farmer-animal relationship score of 3, indicating that the ewes behaved calmly when approached (Figure 2).

Tables 3 and 5 show the mean score of selected animal- and resource-based measurements and production data in 16 farms; 8 farms with the best mean conformation class and 8 farms which had more than 5% of lamb carcases categorised as conformation class P (1–3). The selected farms were qualitatively compared in order to investigate if there

https://doi.org/10.1017/S0962728600002724 Published online by Cambridge University Press

Figure 2



Graphical presentation of the percentage of the human-animal relationship categories 0–3 obtained from 36 sheep farms in three different regions of Norway. Categories of human-animal relationship assessment: 3) Behaved calmly when approached; 2) Some avoidance; 1) Marked avoidance and struggling to escape; and 0) Attempts to escape by jumping out of the pen.

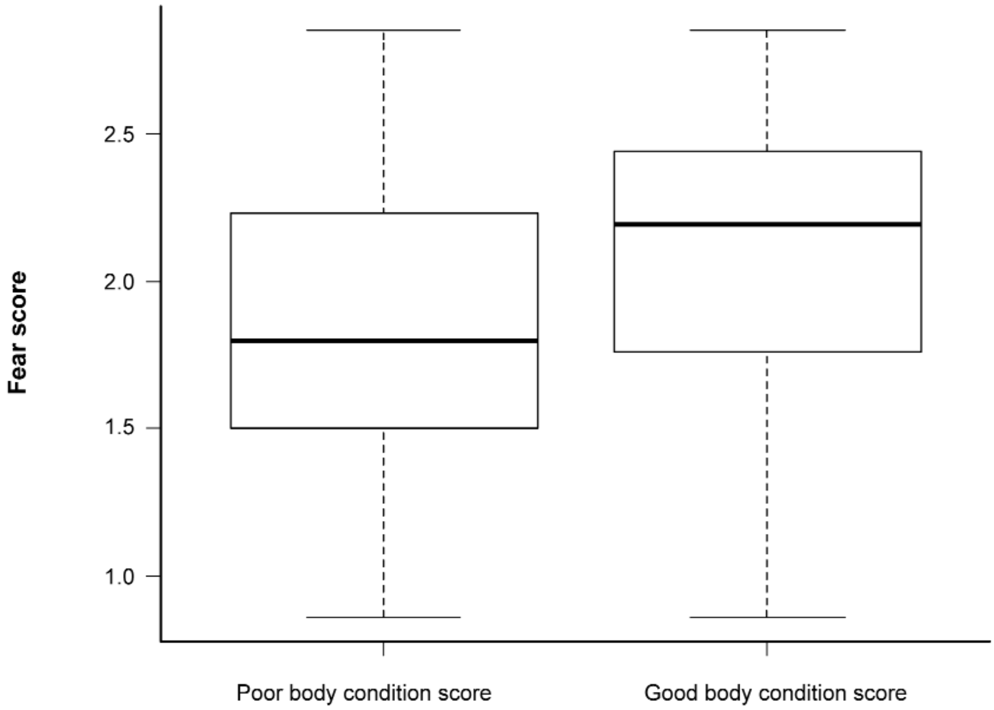
Table 5 Mean score of selected animal[†]- and resource-based measurements and production data in eight farms which had lamb carcases categorised in the best mean conformation class (CC).

	Mean score farmer 1–8									
Variable	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**Mean value 36 farms**	**Recommended values**
BCS[a]	2.9	2.9	2	2.5	2.4	2.7	2.3	3.3	2.6	3[#]
Cleanliness[b]	1.1	2.2	1.4	1.1	1.6	1.9	1.2	2	1.6	1
Skin lesions[c]	1.1	1	1.2	1	1	1.2	1	1	1	1
Skin irritations[d]	1.1	1.2	1.5	1	1.1	1	1	1.5	1.2	1
Callus on carpus[e]	1.3	1.8	1.5	1.6	1.8	2	1.4	1.5	1.5	1
Lighting (lux)	117	175.7	64.5	40.5	51.6	65	56.6	107.2	264	≥ 100 lux[§]
Relative humidity	58.1	75.6	64.1	71.1	81	76.1	87.3	85	66	≤ 70%[§]
Surface temperature (°C)	15.2	7.4	11.6	9.6	8.3	8.6	15	10.9	11	
Draught ($m\ s^{-1}$)	0.5	0.08	0.3	0.11	0.2	0.1	0.4	0.2	0.2	
Ammonia (ppm)	8	4.5	8.3	5.3	5	5.7	10.8	5	5.8	< 10 ppm[§]
CO_2 (ppm)	70	900	57.1	750	625	1,000	110	150	400.2	< 2,500 ppm[§]
Temperature (°C)	16	9.45	12.9	10.3	7.7	7.9	14.1	9.8	11.6	5–25°C
Fear score[‡] (n = 25)		2		2.4	1.7	1.5		2.2	1.9	3
Human-animal relationship*	1	3	2	2	3	1	1	1	1.7	1
CC 1–3 (%)	0	5.6	0	0	0	0.3	1.5	1.6	2.9	
Fat class 1–3 (%)	1.2	7	0.6	1.3	0	0.8	1.5	0.5	3.9	
Slaughter weight (kg)	21.9	22.2	22.3	21.8	23.7	23.1	24.4	23.1	21.4	
Number of lambs slaughtered	162	85	161	80	126	364	68	189	186	

[†] Categories of selected animal-based measurements: [a] Body condition score (BCS), 1: very thin; 2: thin; 3: average; 4: fat; 5: very fat.
[b] Cleanliness, 1: clean; 2: some dirty parts; 3: dirty; 4: very dirty. [c] Skin lesions, 1: no skin lesions; 2: lesions > 1 × 1 cm; 3: ulcerations.
[d] Skin irritations, 1: normal; 2: loss of wool; 3: redness/swelling; 4: parasites or flies.
[e] Callus on carpus, 1: no callus; 2: callus; 3: callus with ulcerated skin.
[‡] Zero indicates the highest level of fear whereas 3 indicates the lowest level.
* Categories of farmer-animal relationship assessment: 3) Behaved calmly when approached; 2) Some avoidance; 1) Marked avoidance and struggling to escape; and 0) Attempts to escape by jumping out of the pen.
[#] Recommended body condition score (BCS) at lambing (Stubbings 2007).
[§] Sevi *et al* 2007.

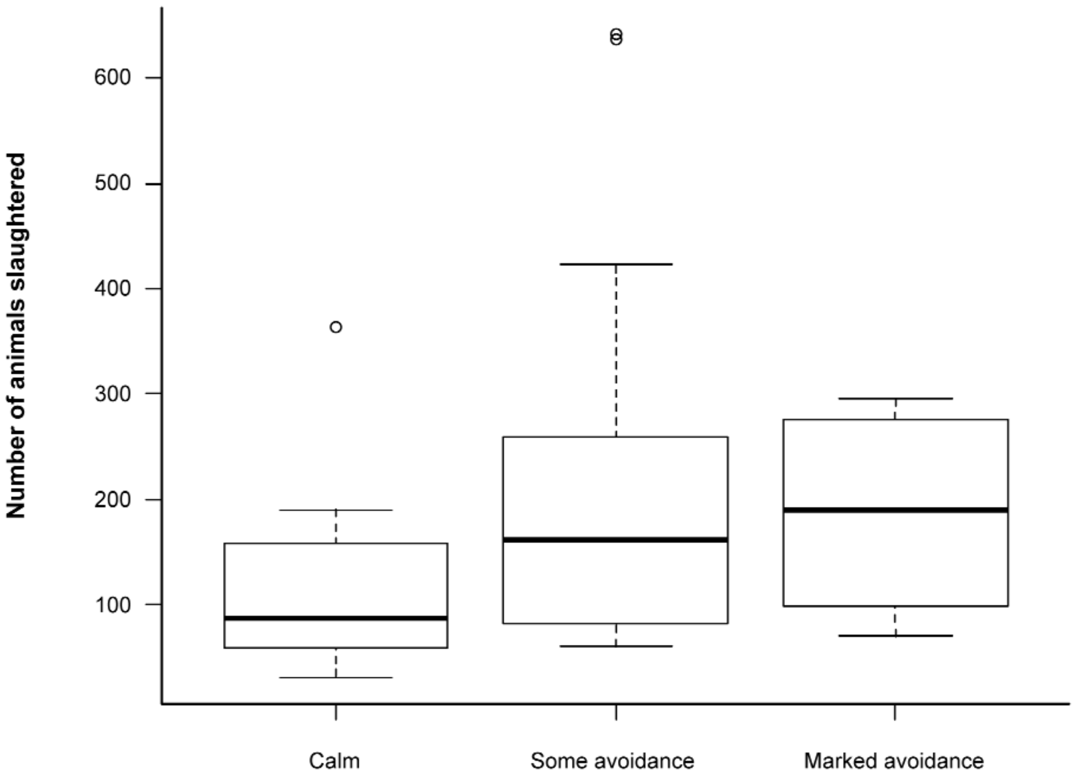
https://doi.org/10.1017/S0962728600002724 Published online by Cambridge University Press

Figure 3



Graphical presentation of the relationship between body condition score (BCS) (n = 250) and fear of humans score. A fear score of zero indicates the highest level of fear whereas a fear score of 3 indicates the lowest level of fear.

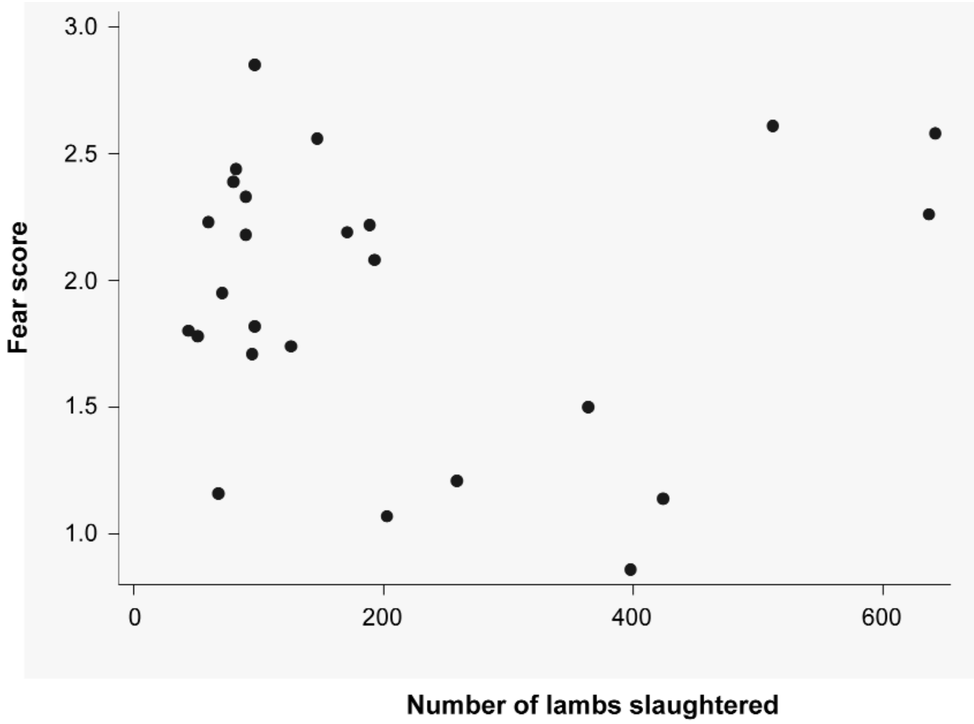
Figure 4



Graphical presentation of the relationship between the farmer-animal relationship scores and the number of lambs slaughtered from individual farms. A higher score depicted a better farmer-animal relationship. Categories of farmer-animal relationship assessment: 3) Behaved calmly when approached; 2) Some avoidance; 1) Marked avoidance and struggling to escape; 0) Attempts to escape by jumping out of the pen.

https://doi.org/10.1017/S0962728600002724 Published online by Cambridge University Press

Figure 5



Graphical presentation of the relationship between the mean fear of human scores and the number of lambs slaughtered from individual farms. A fear score of zero indicates the highest level of fear whereas a fear score of 3 indicates the lowest level.

were patterns in the measurements which could explain the differences in the outcome carcase classification, but no distinctive patterns were detected. A quantitative analysis was also performed, but no correlations were found.

The inter-rater observer agreement was assessed using weighted kappa statistics. The agreement between the two observers was excellent, except for three measures. Body condition score had a kappa value of 0.70. The kappa values for callus on carpus and claws were poor (0.29) and moderate (0.47), respectively.

A significant association was found between the fear score and the body condition score (BCS) of the ewes based on a logistic regression analysis ($P = 0.020$). A lower mean fear score (more fearful ewes) was found to be associated with poor ewe body condition scores (BCS of 1 or 2) (Figure 3).

We observed a tendency for a farmer-animal relationship score of 1 or 2 (more fearful ewes towards the stockperson) to be associated with increased numbers of lambs slaughtered from individual farms, which is indicative of increased flock sizes ($P = 0.10$) based on a logistic regression analysis. Hence, ewes from smaller flocks tended to respond more calmly to the stockperson (Figure 4).

Graphical evaluation of the data suggested a non-linear relationship between the two variables 'fear of human score' and 'number of slaughtered lambs from individual farms', and the plot indicated that the ewes were more fearful in medium-sized flocks (Figure 5). Based on a Kruskal-Wallis equality of populations rank test, we found an association between lower mean fear scores (more fearful ewes) and medium-sized flocks (200–400 lambs sent for slaughter), and a significant difference between small (< 200 slaughtered lambs), medium sized (200–400 slaughtered lambs) and large farms (> 400 slaughtered lambs) ($P = 0.015$).

Discussion

In this study, we developed a protocol for on-farm assessment of sheep welfare, which is presented in Table 1. This protocol is not intended to be definitive, but merely to contribute to the development of a welfare assessment system for sheep and to outline our approach. The protocol was tested on 36 farms, and we investigated how the measured parameters were distributed and if there were any associations between different measurements. The consistency of measurements between observers was also assessed.

Body condition is an important animal-based measurement, and when scored by the system developed by Russel (1984) a ewe should ideally have a body condition score (BCS) of 3 at lambing (Stubbings 2007). The system used in our study was based on a scale of 0 to 5 (Russel 1984). The majority of the ewes had a BCS of 3 (54.7%). However, a relatively large proportion of the ewes had a BCS of 2 (40.6%), which may increase the risk of nutritional stress, disease and low production. Also, maternal bodyweight can have a critical influence on the lambs' birth size (Clarke *et al* 1997). Small birthweight lambs born into a cool, windy environment are especially susceptible to cold stress due to evaporative cooling of foetal fluids on the fleece (Radostits *et al* 1994), thereby increasing the risk of lamb mortality and sub-optimal growth.

To assess the cleanliness of animals is of relevance because the dirt irritates the skin and creates optimal conditions for ectoparasites and other pathogens. Also, dirty animals may indicate a dirty environment or the occurrence of diarrhoea (BWAP 2009). Approximately half of the ewes examined were categorised as dirty to varying degrees. Improvements with regards to cleanliness are important both for animal welfare and the quality and hygienic standards of food production.

Farm animal welfare is affected both by the production system and how the system is managed (Rousing 2003). We assessed the extent to which the animals had sufficient space, proper facilities and company of their own kind, thereby assessing the animals' freedom to express normal behaviour. Resource-based measurements suggested relevant for inclusion in the welfare assessment protocol were temperature, humidity, draught, concentration of gases and amount of light. These environmental factors are easy to measure using objective methods. It is relatively time consuming to measure the size of all pens and trough spaces. We suggest that observing animals while being fed and when lying down could replace these measures in order to assess if the animals have enough space to eat and rest.

Fear is a stressor that may reduce welfare, and the level of fear of the stockperson and an unfamiliar person was therefore assessed. The validated methodology used for the testing of on-farm fear of humans in the present study simply involved marking the animals during feeding (Lankin 1997). This method is fast, simple and inexpensive, and therefore allows quantification of fear of an unfamiliar human using limited time and resources. Concerns have been raised regarding the validity, reliability and feasibility of different methods used to measure animals' response to people in on-farm welfare assessments (de Passillé & Rushen 2005). The extent that animals avoid people will always reflect a mixture of motivations. For instance, an animal's response to people may reflect feeding motivation in addition to fear (de Passillé & Rushen 2005). The ewes tested were fed 2 h prior to testing to standardise this effect. The test used in our study thus measures the approach-avoidance conflict. The motivation to avoid (fear) versus the motivation to eat or explore the human motivates the animals to move in opposite directions. This is the basis of most models of approach-avoidance conflict (unconditioned fear tests) as described in Miller's model (reviewed in Gray 1987). The present test therefore measures the balance between the motivation to approach (hunger and/or curiosity) and to avoid humans (fear).

The farmer-animal relationship test has not previously been validated. However, the method is easy to use and gives an indication of how the stockperson and the animals under his or her care interact. This test should be validated both with regards to a familiar person (farmer) and an unfamiliar person. Forty percent of the ewes tested showed no fear reactions towards an unfamiliar person. Thirty-eight percent of the farms had a farmer-animal relationship score of 3, which indicates that they were calm in the presence of the stockman. We found no correlation between the 'fear of human' and the 'farmer-animal relationship' test, which may be due to the small sample size.

We consider the animal-based parameters to be the most important in our protocol. Due to climatic conditions in Norway, sheep are housed during the winter months. Resource-based parameters were therefore included in the protocol, but this may be of less importance in countries where sheep are not housed to the same extent. The prevalence of several of the animal-based parameters was low in the selected study population. However, this may not be the case in other populations, and we therefore suggest that these parameters should be assessed in a larger population before a conclusion is made as to which parameters should be included in a final protocol. The sample size is relatively small, and the welfare assessment protocol needs to be assessed in a larger number of farms. In addition, the number of ewes examined at each farm could be increased, but this needs to be balanced against time constraints and the feasibility of the protocol.

Assessments from eight farms were qualitatively compared in order to investigate if there were patterns in the measurements which could explain the occurrence of thin lambs at slaughter (Table 3). When several lamb carcases are categorised in class P, this may indicate a welfare problem. Thin lambs can be a sign of internal parasites, diseases or restricted food intake. The latter may be due to decreased access to milk caused for instance by mastitis, poor quality pasture or separation of lambs and ewes on pasture due to predators (Vatn *et al* 2003). Also, assessments from the eight farms with the best mean conformation class were compared with each other (Table 5) and with the other eight farms. No distinctive patterns that could explain the differences were detected. However, the next step would be to select a sample of good and poor performance farms based on carcase classification, and investigate whether any associations between welfare indicators and performance could be detected.

We assessed the inter-rater observer agreement using weighted kappa statistics. In general, the agreement between the two observers was good. However, the selected population was quite homogenous for several variables, which may decrease the conclusiveness of the reliability ratings. The kappa values for callus on carpus and claws were poor and moderate, respectively, which indicate that the scoring systems for these variables require clearer definitions or more in-depth training. The inter-rater observer and test-retest reliability needs to be further assessed in a study with several observers.

We found an association between the fear score and the body condition score (BCS) of ewes. Higher levels of fear (lower mean fear scores as indicated by the fear test) were associated with lower ewe body condition. A relationship between animals' fear of humans and their productivity has been found in numerous studies. For instance, Hemsworth and colleagues (1995) found that 30–50% of the variance between farms in milk production could be explained by the level of fear of humans shown by the cows. It is likely that the BCS of the ewes in this study has been influenced by many factors other than fear, including the quality and amount of feed available. However, fear is a factor that should be taken into consideration by farmers in order both to improve animal welfare and to enhance the animals' productivity.

We also observed a tendency for higher fear (as indicated by the test of farmer-animal relationship score of 1 or 2) to be associated with increased number of lambs slaughtered from individual farms, which is indicative of increased flock sizes. Hence, the ewes appeared to have lower fear and to respond more calmly towards the stockperson in smaller flocks. Lensink *et al* (2000) found that farm size was predictive of the frequency of positive behaviour towards calves by the stockperson, with smaller farms associated with more positive behaviour. It may also be explained in terms of perceived workload. Farmers that own larger farms may have less time to interact with individual animals (English 1991). This result coincides, to some extent, with the result presented in Figure 5, which indicates that the ewes were more fearful in medium-sized flocks. The animals are likely to be more exposed to the stockperson in smaller flocks, thereby habituating to his or her presence during the course of daily management. The less fearful ewes observed in the largest flocks in our study may be due to a more labour-intensive management with several stockpersons, especially during sensitive periods, such as the lambing season, thereby increasing the animals' neutral exposure to humans and giving the stockpersons more time to interact with the ewes. Due to the small sample size, there were only four farms in the medium-sized group and four farms in the large-sized group. This relationship may not exist if the sample size is increased. However, the reported tendency should be investigated further in a study with a greater number of farms included.

Sheep are housed indoors from around October to May in Norway, but the housing period varies considerably in different geographical regions due to varying climate. The developed welfare protocol is intended for use during the lambing season in April–May. Hence, the housing period is long enough to justify the use of welfare indicators based on fear and the human- animal relationship.

The Welfare Quality® protocols are very comprehensive but also relatively time consuming. In this study, we wanted to focus on fewer parameters in order to make the protocol more feasible. The selected parameters should be assessed in a larger study population before a conclusion is made as to which parameters should be included in a final protocol. This study is the first step in the development of a welfare assessment protocol for sheep. Further research is needed to establish the validity of the different welfare indicators, as well as the reliability and feasibility of the selected parameters. A validated protocol for assessment of sheep welfare will be a valuable instrument to help farmers and the industry to assess and improve animal welfare, thereby meeting current societal demands.

Animal welfare implications

In this study, we developed a protocol to assess sheep welfare on-farm. We expect that our contribution will be valuable in the ongoing work to develop a comprehensive and uniform welfare assessment system for sheep, thereby improving sheep welfare.

Acknowledgements

The authors wish to acknowledge the assistance of Ms Hege Gaudernack Tønnesen, Ms Angela Welcome, Ms Stine Løvik Huse, Ms Elise Nagel Alne, Ms Camilla Kielland and Dr Jon Bohlin. The Norwegian Research Council (project number 173313 and 185163), NORTURA and 'Fondet for forskningsavgift på landbruksprodukter' are acknowledged for funding this study.

References

Anonymous 2008 *Annual Report for the Norwegian Sheep Recording System* pp 31. Animalia, Norwegian Meat and Poultry Research Centre: Norway

Blokhuis HJ 2008 International cooperation in animal welfare: the Welfare Quality® Project. *Acta Veterinaria Scandinavica 50*: S10

Blokhuis HJ, Jones RB, Geers R, Miele M and Veissier I 2003 Measuring and monitoring animal welfare: transparency in the food product quality chain. *Animal Welfare 12*: 445-455

Boissy A, Manteuffel G, Jensen MB, Moe RO, Spruijt B, Keeling LJ, Winckler C, Forkman B, Dimitrov I, Langbein J, Bakken M, Veissier I and Aubert A 2007 Assessment of positive emotions in animals to improve their welfare: stress and welfare in farm animals. *Physiology and Behaviour 92*: 375-397

Bøe KE and Simensen E 2003 Simple buildings for sheep: with emphasis on health and welfare. In: *Proceedings of the seminar 'Low-cost housing for ruminants'*. 13-14 October 2003, Sørmarka, Norway

Bristol Welfare Assurance Programme (BWAP) 2009 http://www.vetschool.bris.ac.uk/animalwelfare (Accessed 09.12.09)

Broom DM 1991 Animal welfare concepts and measurement. *Journal of Animal Science 69*: 4167-4176

Clarke L, Yakubu DP and Symonds ME 1997 Influence of maternal bodyweight on size, conformation and survival of newborn lambs. *Reproduction Fertility and Development 9*: 509-514

Commission Regulation (EEC) 1993 *Laying down detailed rules for the Community scale for the classification of carcasses of ovine animals No 461/93.* EC: Brussels, Belgium

Council Regulation (EEC) 1992 *Concerning the Community scale for the classification of carcasses of ovine animals and determining the Community standard quality of fresh or chilled sheep carcasses and extending No 2137/92. Regulation (EEC)* EC: Brussels, Belgium

de Passillé AM and Rushen J 2005 Can we measure human-animal interactions in on-farm animal welfare assessment? Some unresolved issues. *Applied Animal Behaviour Science 92*: 193-209

https://doi.org/10.1017/S0962728600002724 Published online by Cambridge University Press

Dohoo I, Martin W and Stryhn H 2003 *Veterinary Epidemiologic Research. Screening and Diagnostic Tests* pp 91-92. AVC Inc: PEI, Canada

Duncan IJH and Fraser D 1997 *Animal Welfare. Understanding Animal Welfare* pp 19-31. CAB International: Wallingford, UK

English PR 1991 Stockmanship, empathy and pig behaviour. *Pig Veterinary Journal 26*: 56-66

Erhard HW, Fàbrega E, Stanworth G and Elston DA 2004 Assessing dominance in sheep in a competitive situation: level of motivation and test duration. *Applied Animal Behaviour Science 85*: 277-292

Farm Animal Welfare Council 1993 *Report on Priorities for Animal Welfare, Research and Development.* Farm Animal Welfare Council: London, UK

Fisher A and Heal J 2001 *Carcass classification. Livestock Knowledge Transfer, a DEFRA initiative. Beef and Sheep* p 314. ADAS/IGER/University of Bristol: Bristol, UK

Gray JA 1987 *The Psychology of Fear and Stress.* Cambridge University Press: Cambridge, UK

Hemsworth PH, Breuer K, Barnett JL, Coleman GJ and Matthews LR 1995 Behavioural response to humans and the productivity of commercial dairy cows. In: *Proceedings of the 29th International Congress of the International Society for Applied Ethology* pp 175-176. 3-5 August 1995, Exeter, UK

Jones RB 1997 *Fear and Distress. Animal Welfare* pp 75-87. CABI International: Cambridge, UK

Keeling L and Veissier I 2005 Developing a monitoring system to assess welfare quality in cattle, pigs and chickens. In: *Science and Society Improving Animal Welfare. Welfare Quality Conference Proceedings* pp 46-50. 17-18 November 2005, Brussels, Belgium

Kjaernes U and Lavik R 2007 Farm animal welfare and food consumption practices: Results from surveys in seven countries. In: Kjaernes U, Miele M and Roex J (eds) *Attitudes of Consumers, Retailers and Producers to Farm Animal Welfare. Welfare Quality Report No 2*. Cardiff University: Cardiff, UK.

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

Lankin V 1997 Factors of diversity of domestic behaviour in sheep. *Genetics Selection Evolution 29*: 73-92

Lensink BJ, Boissy A and Veissier I 2000 The relationship between farmers' attitude and behaviour towards calves, and productivity of veal units. *Annales de Zootechnie 49*: 313-327

Molony V and Kent JE 1997 Assessment of acute pain in farm animals using behavioural and physiological measurements. *Journal of Animal Science 75*: 266-272

Radostits OM, Blood DC and Gay CC 1994 *Veterinary Medicine: A Textbook of the Diseases of Cattle, Sheep, Pigs, Goats & Horses.* Baillière Tindall: London, UK

Rousing T 2003 *Welfare assessment in dairy cattle herds with loose-housing cubicle systems; development and evaluation of welfare indicators.* PhD Thesis, Danish Institute of Agricultural Sciences, Aarhus, Denmark

Rushen J, Taylor AA and de Passilé AM 1999 Domestic animals' fear of humans and its effect on their welfare. *Applied Animal Behaviour Science 65*: 285-303

Russell A 1984 Body condition scoring of sheep. *In Practice 6*: 91-93

Sevi, A, Casamassima A, Pulina, G and Pazzona A 2007 Factors of welfare reduction in dairy sheep and goats. *Italian Journal of Animal Science 8*: 81-101

Stubbings LA 2007 *Diseases of Sheep. Ewe Management for Reproduction* p 60. Blackwell Publishing: Oxford, UK

Vatn S, Røe M and Nafstad O 2003 *Underfôring av lam, fortsatt et dyrevelferdsproblem?* pp 38-41. Kjøttets tilstand, Fagsenteret for kjøtt. [Title translation: Underfeeding of lambs, still an animal welfare problem?]

Waiblinger S, Boivin X, Pedersen V, Tosi M-V, Janczak AM, Visser EK and Jones RB 2006 Assessing the human-animal relationship in farmed species: A critical review. *Applied Animal Behaviour Science 101*: 185-242.

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

Whay HR 2007 The journey to animal welfare improvement. *Animal Welfare 16*: 117-122

https://doi.org/10.1017/S0962728600002724 Published online by Cambridge University Press