On-farm qualitative behaviour assessment in sheep: repeated measurements across time, and association with physical indicators of flock health and welfare

Phythian, CJ; Michalopoulou, E; Cripps, PJ; Duncan, JS; Wemelsfelder, F

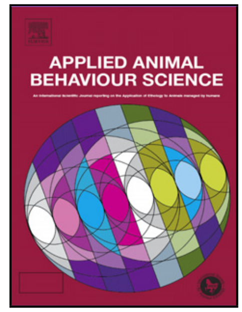
Published in:
Applied Animal Behaviour Science

DOI:
10.1016/j.applanim.2015.11.013

First published: 28/11/2015

Document Version
Peer reviewed version

Accepted Manuscript



Title: On-farm qualitative behaviour assessment in sheep: repeated measurements across time, and association with physical indicators of flock health and welfare

Author: Clare J. Phythian Eleni. Michalopoulou Peter J. Cripps Jennifer S. Duncan Françoise. Wemelsfelder

PII: S0168-1591(15)00317-2
DOI: http://dx.doi.org/doi:10.1016/j.applanim.2015.11.013
Reference: APPLAN 4162

To appear in: *APPLAN*

Received date: 17-5-2014
Revised date: 7-10-2015
Accepted date: 14-11-2015

Please cite this article as: Phythian, C.J., Michalopoulou, Ei., Cripps, P.J., Duncan, J.S., Wemelsfelder, Fe.,On-farm qualitative behaviour assessment in sheep: repeated measurements across time, and association with physical indicators of flock health and welfare, *Applied Animal Behaviour Science* (2015), http://dx.doi.org/10.1016/j.applanim.2015.11.013

Highlights

- QBA assessments of sheep flocks over a year were compared to flock health measures
- Two dimensions of sheep expression, summarised as ‘mood’ and ‘responsiveness’
- Flock scores on both dimensions showed high consistency across the year
- Flock ‘mood’ scores correlated to flock lameness and ‘dull physical demeanour’
- Results support QBA as a meaningful, complementary sheep flock welfare indicator

On-farm qualitative behaviour assessment in sheep: repeated measurements across time, and association with physical indicators of flock health and welfare

Phythian, Clare J.[1a], Michalopoulou, Eleni.[2], Cripps, Peter J.[2], Duncan, Jennifer S.[2], and Wemelsfelder, Françoise.[3]

[1] Department of Epidemiology and Population Health, University of Liverpool, Leahurst, CH64 7TE, United Kingdom

[2] School of Veterinary Science, University of Liverpool, Leahurst CH64 7TE, United Kingdom

[3] Animal and Veterinary Sciences Group, SRUC, Roslin Institute Building, Bush Estate, EH25 9RG, United Kingdom

Corresponding author: clare.phythian@nmbu.no

[a] Current address: Section for Small Ruminant Research, Department of Production Clinical Science, NMBU, 4325 Sandnes, Norway

Telephone: +47 51603510

Abstract

Qualitative Behavioural Assessment (QBA) is a 'whole-animal' methodology that assesses the expressive qualities of animal behaviour using terms such as 'tense', 'relaxed', 'anxious', and 'content'. The reliability and validity of QBA as an indicator for on-farm welfare assessment in pigs, cattle, poultry and sheep has been examined in a number of ways. However, the use of QBA on farms over longer periods of time has not yet been examined. The aim of this study was to investigate whether and how on-farm QBA of sheep varies over the different seasons of the year, and whether it is associated with physical measures of sheep health and welfare such as lameness. A trained assessor visited each of 12 farms six times within a one year period at two month intervals, and made group level assessments of approximately 100 sheep selected *ad hoc* (assuming homogeneity within the flock). The sheep flocks were assessed with a list of twelve QBA descriptive terms previously developed for sheep. Following QBA, the same sheep were also assessed with seven physical indicators of health and welfare ('dull physical demeanour', lameness, breech and abdominal soiling, pruritis, wool loss, and coughing). QBA scores from all visits were analysed together, and also in combination with the physical measures, with Principal Component Analysis (PCA - correlation matrix, no rotation). The effect of visit on PCA flock scores was analysed with random-effects multiple linear regression models. The association between PCA flock scores and physical measures was investigated using Spearman rank correlation (rS), and the correlation of flock rankings across visits was examined with Kendall Coefficient of Concordance. PCA distinguished two main dimensions of sheep expression: PC1 (47% variation) ranging from content/relaxed/thriving to distressed/dull/dejected (summarised as 'mood') and PC2 (21%), which ranged from anxious/agitated/responsive to relaxed/dejected/dull (summarised as 'responsiveness'). No significant effect of visit on PC1 scores was found (p=0.155), and PC1 flock scores correlated at W=0.84 (p<0.001) across the 6 visits, indicating high consistency of characterisations of individual flock mood over the year. However there was an effect of visit on PC2 scores (p<0.001), and PC2 flock scores were correlated at W=0.60 (p<0.001) across visits, indicating that the presence of young lambs may have had a consistently relaxing effect on flocks. There was also an effect of visit period on lameness (p=0.025), and on breech (p<0.001) and abdominal (p=0.0048) soiling. With the exception of lameness and breech and abdominal soiling, the physical indicators were observed at a low prevalence (<2%) across the study farms. The highest lameness levels were observed during the winter period (mean 17.86%, 95% CI 7.83 – 27.90) whilst breech soiling was highest in spring (mean 23.83%, 95% CI 11.86 – 35.81). An effect of farm type was found on lameness scores (p=0.0176) and an effect of flock size on abdominal soiling scores (p=0.025). PC1 'mood' scores were negatively correlated to the proportion of lame sheep (n=72; rS=-0.72, p<0.001), and to the proportion of animals with dull physical demeanour (rS=-0.70, p<0.001), while PC2 'responsiveness' scores showed a weak correlation with breech soiling (rS=0.42, p<0.001). In summary, these results suggest that QBA has the potential to serve as a sensitive, meaningful indicator for on-farm welfare assessment in sheep.

Keywords: Qualitative Behaviour Assessment; animal welfare; consistency; animal-based outcomes; sheep.

1. Introduction

Qualitative Behavioural Assessment (QBA) is a 'whole-animal' methodology that assesses the expressive qualities of animal behaviour using terms such as 'tense', 'relaxed', 'anxious', and 'content'. Thus it addresses an animal's 'body language', including both negative and positive aspects of well-being, and has the potential to integrate and help interpret specific clinical measures of physical and psychological health (Napolitano *et al.*, 2009; Wemelsfelder & Lawrence, 2001; Wiseman-Orr *et al.*, 2006). This methodology has been applied to assess animals on-farm and during transport both individually and at group-level, with different livestock species such as pigs, cattle, poultry and sheep (e.g. Bassler *et al*., 2013; Rousing and Wemelsfelder, 2006; Stockman *et al.*, 2011; Temple *et al.*, 2011; Wickham *et al.*, 2012). Generally good levels of inter-observer reliability (but not always, see Bokkers *et al.*, 2012), meaningful associations with other measures (but not always, see Andreasen *et al*., 2013), as well as short assessment times, suggest this method has the potential to be an effective welfare indicator that can be readily applied in the field.

In common with other global pasture-based production systems, sheep managed under British farming systems spend a considerable part of the production cycle outdoors at pasture being kept in specific management groups. Therefore, groups of sheep often require gathering and handling to facilitate close inspection and assessment of the health and welfare of both the individual sheep and the flock. Since disturbance by humans, dogs and handling can alter ovine behavioural expression (Boivin *et al.*, 2000, Le Neindre *et al.*, 1996) and mask painful conditions (Fitzpatrick *et al.*, 2006), it is possible that some sheep with welfare issues may be missed when gathered for closer examination. Furthermore, the practicalities of assessment need to be considered for different management systems. The gathering and handling of extensively-managed sheep and those managed over multiple locations can be time and labour consuming and also may not be appropriate at certain periods of the production calendar, for example, when ewes have young lambs at foot or during the mating period. Therefore, a welfare indicator that does not involve major disturbance, requires few resources, and offers valid information on the health and wellbeing of groups of animals, could offer clear benefits for sheep, producers and assessors.

One major concern in the development of on-farm welfare assessment protocols is the challenge of interpreting fluctuations shown by welfare indicators across time. Such fluctuations may be part of normal day-to-day or seasonal variations in welfare, may reflect more serious deviations of basic welfare, or could reflect the effects of varying times and contexts on repeat assessments. Thus, if repeated assessments of the same farm do not show similar levels of animal demeanour, it is difficult to know whether this difference reflects normal baseline variation, a welfare problem, or a problem of intra-observer reliability (Temple *et al.*, 2013). The aim of this study was to apply QBA to the repeated assessment of sheep at flock level in a one-year longitudinal study, to investigate whether and how the sheep's expressive demeanour would be perceived by an experienced assessor to vary across 6 visits at two-monthly intervals. To evaluate these assessments against other welfare indicators, seven physical measures of sheep health and welfare were also examined.

2. Materials and methods

2.1 Design of longitudinal study

A longitudinal on-farm study performed during the period of May 2009 – April 2010 was conducted on twelve farms, located in North-West England and North Wales, which had previously participated in a sheep welfare research project. Farms were selected according to their location, farm type and owner's informed consent to participate. Selection provided a sample of eleven commercial flocks and one small-holding, including farms from hill, upland and lowland areas (for details see Table 1). At each visit, each farm was asked to provide a sample of 70-100 sheep that were selected *ad hoc* by the farmer and left undisturbed for assessment. This sample size was not related to a farm's flock size, but was based on previous experience of the assessor regarding the feasibility of completing the protocol of qualitative and quantitative assessments within the time limits of a day visit. The exact numbers of sheep selected at each farm for each visit were recorded. The study was approved by the University of Liverpool Ethics Committee (ethical review reference number RETH000287).

During the one year study, flocks were repeatedly assessed by one sheep veterinary surgeon who performed all QBA and physical indicator assessments on all farms throughout the study. Repeated sampling of twelve sheep flocks over 6 visits spread out over one year produced 72 on-farm assessments. Flocks were visited at an interval of approximately 60 days, to coincide with key periods in the sheep production cycle (Table 2). At each visit, the selected group of sheep was firstly assessed using twelve QBA descriptors (relaxed, dejected, thriving, agitated, responsive, dull, content, anxious, bright, tense, vigorous and distressed), which had previously been developed and tested for inter-observer reliability by Phythian *et al.* (2013a). Due to their integrative, qualitative nature, it is impossible to define QBA terms in precise physical terms such as is done for conventional ethograms (however very recently QBA studies have begun to provide brief qualitative characterisations of individual terms to enhance observer agreement). Detailed instructions for how to score QBA terms were developed for the Welfare Quality® protocols for cattle, pigs and poultry (Welfare Quality®, 2009), including careful reflection on, and, where more than one assessor are involved, discussion of, the meaning of individual terms. These instructions were followed in the present study.

The assessor quietly approached the sample group and performed assessments from a distance by standing at the boundary of a field, or several metres from groups of housed animals. The exact sizes of fields and assessment areas were not measured, but a number of observation points was selected according to the relative size of the field and sample group, after which a 5 minute period was allowed to let sheep get accustomed to the presence of the assessor. The mean number of sheep assessed in any one group was 77, and ranged from a minimum group size of 24, which represented all the flock of a small-holding farm, to a maximum group size of 137 animals on a commercial farm. Minimal disturbances of the sheep by assessor movements, particularly in situations where scrutiny of individual animals was difficult, were found to be helpful and considered acceptable. The observer then spent 5 minutes at each of the observation points, visually scanning the designated observation area to assess the entire sample group of sheep. When observations were completed, the groups' predominant behavioural expressions were scored on each of the QBA terms along a visual analogue scale (VAS) of 125 mm length, labelled from 'zero' to 'maximum' expression. This entire process of QBA assessment, of up to 120 sheep, took on average about 30 minutes per farm.

Following completion of QBA, seven additional physical indicators of sheep health and welfare were assessed at group-level by the same assessor. Whilst these physical measures were taken in the same observation area as QBA, the exact observation points from which they were made differed. The group of sheep was briefly observed at a distance for five minutes, and then the assessor entered the assessment area to count the number of animals observed to be affected by the following physical indicators (as described in Phythian et al., 2012): coughing (defined as observation of one, or a combination, of the following signs: paroxysmal coughing, and respiratory distress including abdominal effort associated with breathing or wheezing), lameness (any, or a combination, of the following signs: 'nodding' of head in unison with short stride, grazing on knees, uneven gait, arching of back during locomotion, non-weight bearing on affected limb when standing, extreme difficulty rising, and reluctance to move once standing, as described in Kaler *et al*., 2009), breech soiling (discrete/solid plaques or more diffusely soiled areas of contamination by faecal matter, mud or soil of the perineum and/or tailhead, and/or superficial gluteal region, and/or caudal aspect of the hindlimb(s) as far as the hock), abdominal soiling (discrete/solid plaques or more diffusely soiled areas of contamination by faecal matter, and/or mud, and/or soil over the ventral abdomen), pruritis (one, or a combination, of the following signs: rubbing or scratching against walls/posts/fences/other objects, restlessness, stamping of feet, biting and nibbling of own body), wool loss (observation of small discrete areas extending to diffuse areas of fleece loss), and 'dull physical demeanour' (defined as "an animal with lowered head carriage, showing behavioural separation from the rest of the group, and unresponsive to the presence of other sheep or the observer").

2.2 Statistical analysis

QBA data for each farm were recorded by measuring the distance in millimetres between the zero point of the VAS scale and the mark on the line made on the scale for each term, to provide a value between 0 and 125. For physical health and welfare indicators, the percentage (%) of sheep in a group showing signs of coughing, wool loss, pruritis, lameness, breech and abdominal soiling, and 'dull physical demeanour', was calculated for each farm assessment.

QBA data recorded over the 6 visits were analysed together using Principal Component Analysis (PCA – correlation matrix, no rotation) in Minitab version 16 (Minitab, Inc, State College, PA). PCA identifies the least number of components that explain most of the variance in the data (Jolliffe, 2002). QBA PC1 and PC2 accounted for a cumulative variance $\geq$ 68 %, hence two components were retained in subsequent analyses. The correlation between each QBA term and PC1 and PC2 is contained in the loading values, which reflect the weighting of each term within each component. In addition, a combined PCA analysis (correlation matrix, no rotation) was performed by analysing data on all 19 variables (12 QBA terms and 7 physical indicators) gathered over the 6 visits.

To investigate whether PC1 and PC2 scores for the 12 flocks differed in ranking over the 6 visits, Kendall coefficient of concordance (W) was calculated (n=12). Correlations of PC1 and PC2 flock scores with the outcomes of physical health and welfare indicators were examined using Spearman's rank correlation coefficient (rS, n=72), and the distributions of QBA PC1, PC2 and physical indicator scores for each farm over the 6 visits were examined graphically.

To investigate whether there was a significant effect of visit period on QBA and physical indicator scores, mixed effects linear regression models were fitted in Stata version 13.1 (StataCorp LP, College Station, TX). Visit period (n=6; Table 1) was included as a fixed effect, and farm identity as a random effect. Farm type (categorised for the purposes of analysis as 1. lowland or 2. hill and upland) and flock size (categorised as ≤100, 101 – 350, 351 – 650, 651 – 950 and 951 – 1250 sheep) were also included as covariates in the mixed effects models. Visit period (1 to 6) was included as a repeated measure within-farm. To ensure the robustness of regression models, an auto-regression correlation was also fitted in the order of 1. Models examined the effect of visit period on both the QBA PC1 and PC2 scores, and on the combined QBA/physical indicator PC1 and PC2 scores. However, due to paucity of data for several physical indicators only those indicators observed at a prevalence >2% (lameness, and breech and abdominal soiling) were included. The models' outcomes were described using coefficient β (indicating the magnitude of the effect), a 95% confidence interval (CI), and Wald p-values (Long and Freese, 2006). To assess the effect of visit period, the baseline (β=0) for comparison of coefficient values for each visit period was set as visit 1 (May-June 2009). Lowland farms and flocks with less than 100 sheep were set as the baseline values (β=0) for comparing the effects of farm type and flock size respectively.

3. Results

A total of 5740 sheep (aged > 1 year) and lambs (aged > 12 weeks) were assessed, using QBA and 7 physical indicators of sheep health and welfare. Over the six visits the total number of sheep presented for assessment on each farm were: farm 1 = 481, farm 2 = 552, farm 3 = 216, farm 4 = 447, farm 5 = 428, farm 6 = 567, farm 7 = 439, farm 8 = 525, farm 9 = 553, farm 10 = 529, farm 11 = 471 and farm 12 = 532. The total number of sheep assessed per visit varied: visit 1: n = 1182, visit 2; n = 1133, visit 3, n = 990; visit 4, n= 780; visit 5, n = 709; and visit 6, n = 946.

3.1 QBA and physical health indicator outcomes

PCA identified two principal components (PC) which together explained 68% of the variation between farms (Fig. 1). PC1 (47% variation) ranged from ‘content/relaxed/thriving’ to ‘distressed/dull/dejected’ (summarised as ‘mood’), while PC2 (21%) ranged from ‘anxious/agitated/responsive’ to ‘relaxed/dejected/dull’ (summarised as ‘responsiveness’). The proportion of sheep observed with signs of each physical indicator varied between individual farms. Across the entire study period, at the level of the individual farm, ‘dull physical demeanour’ ranged from 0 (minimum) to 15% (maximum), coughing from 0% to 38.55%, wool loss from 0% to 11.54%, pruritis 0 to 2.88%, lameness 1.23% to 61.86%, whilst breech and abdominal soiling ranged from 0 to 59.68% and 0 to 100% respectively. In addition, as can be seen from Table 3, there was seasonal variation between different visits in the mean proportion of affected sheep for all study farms.

3.2 Effects of visit period

Overall, regression modelling identified no significant effect of visit periods 1-6 on PC1 flock scores when QBA scores were evaluated independently ($p=0.155$), nor when analysed together with physical indicator outcomes ($p=0.1982$). This result indicates that the perceived mood of sheep flocks was relatively stable across the year. By contrast, there was a significant effect of visit period on PC2 scores for both the independent QBA PC2 scores ($p<0.001$), and the combined QBA/physical indicator PC2 scores ($p<0.001$). More detailed results for the random-effects models are shown in Table 4, which show that the lowest β coefficient values for PC2 scores were associated with visit 1 (May/June 2009) and visit 6 (March/April 2010), indicating that flocks appeared relatively more relaxed and less agitated over the lambing and post-lambing period than at other times of year. The significant correlations between the rankings of flocks on PC1 and PC2 across the 6 visits indicate that the relative characterisations of flock expression on the two QBA dimensions did not significantly change over the year (PC1: $W=0.84$, $p<0.001$; PC2: $W=0.60$, $p<0.001$). As regards physical indicators, there was an effect of visit period on lameness and breech and abdominal soiling scores. The highest level of lameness (17.86%) was recorded in the winter months (visit 5), breech soiling (23.83%) was highest at visit 1 in the spring period (Table 4), and the highest levels of abdominal soiling (20.17%) occurred during the autumn/winter period (visit 4).

3.3 Effects of farm type and flock size

An effect of farm type was found on QBA PC1 scores in which hill/upland flocks received higher PC1 scores (β 2.47, 95% CI 0.54 – 4.38, $p=0.017$), and were thus perceived as more 'content/relaxed/thriving', compared to the lowland flocks in this sample (β -1.21, 95% CI -2.57–0.14). There was also an effect of farm type on lameness scores, indicating that hill/upland flocks showed lower levels of lameness (β -9.70, 95% CI -17.72 – -1.68, $p=0.0176$) than lowland flocks (β 18.32, 95% CI 12.66 – 23.99). There was one effect of flock size: larger flocks with 951–1250 sheep showed higher levels of abdominal soiling (β 28.28, 95% CI 1.94 – 29.17 $p=0.025$) than flocks with less than 100 sheep (β 0, 95% CI -11.12 – 11.11). No significant interaction effects were found.

3.4 Associations between QBA and physical health indicators

Fig. 2 shows PC1 and PC2 of the combined PCA of QBA and physical indicator scores. This graph illustrates a close alignment of the negative end of PC1 (distressed/dull/dejected) with the prevalences of lameness ($r_S=-0.72$, $p<0.001$), and 'dull physical demeanour' ($r_S=-0.70$, $p<0.001$).

This association is supported by significant correlations of lameness with individual QBA terms ‘distressed’ (r_S=0.50, p<0.001), ‘dull’ (r_S=0.57, p<0.001), and ‘dejected’ (r_S=0.57, p<0.001), and by a correlation of ‘dull physical demeanour’ with individual QBA terms ‘distressed’ (r_S=0.70, p<0.001), ‘dull’ (r_S=0.74, p<0.001), and ‘dejected’ (r_S=0.66, p<0.001). In addition there was a weak but significant correlation between the negative end of PC2 (relaxed/dejected/dull) and breech soiling (r_S=0.42, p<0.001).

Fig. 3 presents a visual image of the distributions of QBA PC1 and PC2 scores, and lameness and breech soiling percentages, for each of the 12 farms across all 6 visits. This overview allows closer investigation of the extent to which the PC scores of different farms remained stable, or shifted up or down the expressive dimensions. Thus some farms (e.g. farm 3 in Fig.3) did not vary much in their PC1 and PC2 positions over the year, whereas other farms showed more variation over time (e.g. farm 4). The majority of flocks stayed located on either the positive or negative side of PC1, and thus appeared to be quite consistent in general mood (as supported by a high Kendall W for PC1 of 0.84). On PC2 there was more variation for some farms (e.g. farms 4 and 5), reflected in a somewhat lower, but still significant, Kendall W value (0.60). The effect of farm type on QBA PC1 scores and lameness scores reported above can also be seen in this graph: hill farm flocks (farms 3, 9, 10, 11, and 12) were assessed as relatively content, relaxed and thriving compared to lowland flocks (farms 1, 2, 4, 5, 6, and 7), and also showed consistently lower levels of lameness across visits than lowland flocks. Particularly in flocks 1 and 4 the association between low PC1 scores/negative mood and high levels of lameness is evident.

4. Discussion

This study applied qualitative behaviour assessment (QBA) to the on-farm assessment of sheep welfare, using a list of QBA terms developed for sheep by Phythian *et al.*, (2013a). The emphasis in this study was on repeated QBA assessment of sheep flocks across time, over 6 visits at two-monthly intervals, on 12 hill and lowland farms in the UK. The study's aims were to assess whether and how QBA, applied by an experienced assessor, was capable of detecting differences in sheep behavioural expression over time, and was associated with physical health measures taken at the same time points. Multivariate analysis identified two main dimensions of sheep expression: PC1, ranging from content/relaxed/thriving to distressed/dull/dejected (summarised as 'mood'), and PC2, ranging from anxious/agitated/responsive to relaxed/dejected/dull (summarised as 'responsiveness'). These dimensions correspond well to those found (dim1: content/relaxed/bright to distressed/dejected/tense; dim2: agitated/responsive/anxious to dull/dejected/relaxed) in a study by Phythian *et al*. (2013a), in which 13 veterinary and farm assurance assessors provided QBA, using the same terms as the current study, of 12 video clips showing sheep in varying indoor and outdoor situations and housing conditions. This convergence supports the relevance of these dimensions for characterising sheep expressions in varying on-farm conditions.

Previous QBA studies (e.g. Rutherford *et al*., 2012) have frequently found two main dimensions of behavioural expression, where the first dimension corresponds to a distinction between positive and negative experience, and the second dimension appears to distinguish between low and high levels of arousal/activation in these experiences. The dimensions identified in the present study concur with this pattern, with positive and negative descriptors aligning on opposite sides of PC1, and high-arousal/activation terms (e.g.: agitated, responsive) placed on the opposite side to low-arousal terms (e.g.: relaxed, dull) on PC2. The four quadrants thus formed appear to fit in well with the integrative functional framework for emotion and mood proposed by Mendl *et al*. (2010), supporting our summarising labels of PC1 and PC2 term-loadings as 'mood' and 'responsiveness'. This is not the place to discuss the relationship of QBA with cognitive or motivational theoretical frameworks of emotion, and we do not wish to suggest that QBA studies actually measure any cognitive or motivational states inferred by such frameworks; however QBA assessments of animals' dynamic expressive demeanour do generally appear to be compatible with these frameworks (Boissy *et al*., 2007; Mellor, 2012).

4.1 Detecting differences over time

This study's outcomes indicate that on average, PC1 flock scores (whether QBA scores were evaluated independently or combined with physical indicator outcomes) did not differ across the 6 visits made in the course of a year, while the rankings of flocks on PC1 were highly correlated across visits. Together these findings indicate a high consistency of QBA assessments of the flocks' mood over the course of the year. To our knowledge, only one previous on-farm study with pigs has investigated repeated application of QBA over a longer time period, finding moderate correlation (rS=0.50) between the PC1 scores generated by two visits (1 year apart) to the same 15 intensive pig farms (Temple *et al.*, 2013). The authors of the pig study rightly note that when using one assessor, it is not possible to tell whether lack of good correlation is due to poor intra-observer reliability or to a genuine change in the animals' state. Weaker correlations may also reflect a lack of sufficient between-farm variation in the same production system. The present sheep study assessed the same farms 6 rather than 2 times a year, and included farms from different production systems, allowing closer examination of variation between and within farms. The high consistency found in PC1 farm scores across visits, both within and between farms, suggests that the flocks' mood remained relatively stable on most farms across the study period.

PC2 flock scores (responsiveness) did show significant variation across visits, with the lowest scores (i.e. most relaxed/dejected/dull) recorded post-lambing in May/June (visit 1), and at the 2010 lambing season (visit 6). An explanation for this might be that the presence of lambs less than 12 weeks old had a relaxing effect on sheep, due to physiological changes associated with maternal bonding behaviours such as licking and grooming of lambs (Dwyer *et al.,* 2008). It is also conceivable that variations in sheep responsivity reflected arbitrary differences in how sheep were selected for assessment by farmers. However, there were no signs of deliberate bias in how farmers selected sample animals. Moreover, as management practices tend to affect the whole flock rather than specific animals, repeated assessment of different groups of sheep from the same farm was considered to provide a representative sample of the flock. This view is supported by our finding that the rankings of individual sheep flocks on PC2, like those on PC1, were significantly correlated across the 6 visits.

Levels of lameness and breech and abdominal soiling varied over the year. The mean lameness level of 7.23% observed during the first visit period (May/June 2009) is close to the mean lameness estimate of 7.10% previously identified during a cross-sectional study on 40 English and Welsh sheep farms (Phythian *et al*., 2013b). However, at 13.39% the mean level of lameness for all flocks observed across the whole study period is considerably higher than previously reported, but does concur with King and Green's (2011) conclusions regarding the true prevalence of flock lameness in England. The high lameness levels observed in the current study may reflect the convenience-based approach to farm sampling, or be due to specific disease outbreaks in these flocks. High prevalence of lameness appears to coincide with periods of gathering, handling, and housing of animals, which favour the transmission of infectious causes of ovine lameness such as footrot (Raadsma and Egerton, 2013). Indeed in the current study the highest levels of lameness were recorded during autumn, coinciding with the mating period, and in winter, when most sheep flocks were housed. Follow-up veterinary examination after assessments were completed indicated that farms with high lameness prevalence showed high levels of infectious footrot, and problems with controlling an outbreak of contagious ovine digital dermatitis were thought to explain the very high (up to 61.86%) lameness levels recorded on farm 1. The increased level of breech soiling at the spring visit may be due to changes in nutrition, and to the greater parasite challenges of the spring grazing season. On the other hand, the higher level of abdominal soiling observed in the winter period may relate to the wet weather conditions experienced by some sheep flocks that had not yet been winter housed.

4.2 Effects of farm type and flock size

There were effects of farm type on both QBA and physical indicator scores. Flocks on hill/upland farms were assessed as more content, relaxed and thriving (PC1) than lowland flocks. This may reflect a difference between extensive and intensive management practices, but could also reflect the effects on welfare of differences in physical health. Indeed, we found that hill farm flocks showed consistently lower levels of lameness than lowland farms, and there was a significant negative correlation between PC1 mood scores and lameness (for further discussion of this association see below). Thus the hill farms selected for this study appeared to have a higher health and welfare status than the lowland farms, however, the study's non-random sampling approach, and the small sample of 12 farms, prevent a more general interpretation of these findings in terms of sheep welfare in different production systems. The difficulty of establishing larger samples with sufficient farm-level variation is a common feature of applied on-farm research (Andreasen *et al.*, 2013).

There was no significant effect of farm type or flock size on the sheep's responsiveness (PC2); however visual inspection of PC2 scores (Fig. 3) suggests some meaningful variation at individual farm level: sheep on a small-holding in a hill area (farm 3), which were regularly handled and petted, appeared consistently more relaxed than sheep on hill farms under more extensive management (farms 9-12). This difference may reflect the positive effect of handling on welfare (Boivin *et al.,* 2003), and suggests that QBA can generate expressive dimensions on which management practices can be meaningfully evaluated. However further study with larger sample sizes is required to support these findings. There was one effect of flock size on physical health, with large flocks (950-1250 sheep) showing a higher prevalence of abdominal soiling than sheep in small flocks (less than 100 sheep), which may reflect the greater exposure of large extensively managed commercial flocks to environmental and climatic conditions. However here too, further study with larger sample sizes is required.

4.3 Associations between QBA and physical health indicators

Of the seven physical indicators of sheep health and welfare recorded, only the proportion of lame sheep on a farm, and the proportion of sheep recorded with 'dull physical demeanour', correlated significantly with PC1 mood scores, indicating that flocks with high levels of lameness and dull physical postures were perceived as more distressed/dull/dejected than other flocks. Indeed visual inspection of individual farm scores (Fig. 3) indicates that farms with consistently low mood scores (e.g. farms 1 and 4) showed high levels of lameness. Lameness is a key welfare issue for sheep, and dull physical demeanour is an attribute commonly assessed by stock-people and veterinarians to detect sickness and disease. The significant association of QBA expressive dimensions with these health measures supports that QBA addressed important aspects of sheep welfare, and provided complementary information to help interpret the wider welfare impact of these health problems. Lameness is associated with pain, however the present study described lame sheep as more distressed, dull, and dejected than non-lame sheep, suggesting that lameness and the pain underlying it also had a more generally deleterious effect on the sheep's emotional state. The same can be said for dull physical demeanour; this is a specific physical measure usually associated with sickness and pain, however the present study suggests these clinical signs also had a more generally deleterious emotional effect.

One could argue that the QBA terms 'dull', 'dejected', and 'responsive' and the physical indicator 'dull physical demeanour' can hardly be considered independent measures, and may have been subject to what Greenwald *et al*., (1986) call 'theory confirmation bias'. However the two measures were embedded in very different assessment and scoring procedures, and were not taken closely together. QBA was always performed right at the start of the assessment, ensuring independence of any physical measurement afterwards. Moreover, PC1 mood scores were not the result of direct measurement, but the outcome of a multivariate analysis, creating more analytical distance between the two types of measure. Their significant correlation was thus not pre-given, but can still be considered a meaningful outcome. A good degree of association between different types of measure confers an aspect of internal validity known as convergent validity (Abramson and Abramson, 2008). Yet to confer validity this association does not have to be maximal, in that the different types of measure address the animal welfare construct from different angles: that of psychological well-being (QBA), and physical health (lameness, dull physical demeanour).

Coughing, wool-loss, abdominal soiling, and skin irritation were observed too infrequently to enable meaningful correlation with QBA. Breech soiling did occur frequently, and there was a weak but significant correlation between breech soiling and PC2, indicating that flocks that were more relaxed were also more soiled. The reasons for this association are likely to be complex, but may be related to a co-variance in time: both measures increased significantly in spring. As discussed above, sheep may have become more relaxed after lambing, and this may have affected their behaviour in a way (e.g. increased lying) that led to more breech soiling in wet weather conditions. Breech soiling can also be influenced by seasonal dietary changes, or indicate the presence of endo-parasitism or blowfly myiasis (French *et al*., 1994), which may have caused the sheep to slow down, but on these farms it did not appear to affect their welfare such that they became dull and dejected, as was the case with lameness.

Thus, combining QBA with other indicators is likely to provide a fuller, more complex picture of animal health and welfare (Wemelsfelder & Farish, 2004; Wemelsfelder & Mullan, 2014). Through patterns of correlation, QBA can help to interpret health and behaviour measures in terms of an animal's well-being, as has for example been reported for sheep (Wickham *et al*., 2012), horses (Minero *et al*., 2009), and pigs (Rutherford *et al*., 2012). An on-farm assessment study of 43 Danish dairy cattle farms found no meaningful correlations between QBA and measures of the

Welfare Quality® assessment protocol (Andreasen *et al*., 2013). However, as discussed above for the Temple *et al.* (2013) on-farm pig study, this may have at least partly been due to a lack of sufficient between-farm variation in the study's sample. A risk factor analysis of on-farm assessments of 89 commercial broiler farms in various EU countries reported some meaningful associations between QBA and other measures (Bassler *et al*., 2013). The present study included farms from different sheep production systems; that despite its small sample size it found significant associations between QBA and physical health measures is encouraging, and should stimulate more research.

4.4 Methodological considerations

For reasons of feasibility and cost, all assessments in this study were performed by the same assessor. It could be argued therefore that the repeated farm assessments were not independent, and may have artificially inflated the stability of PC1 mood scores across the year. Indeed, as noted by Temple *et al.* (2013), when using a single assessor it is not possible to distinguish observer from farm effects, other than by consideration of a study's larger context and totality of findings. However, PC2 responsiveness scores did fluctuate across visits, which suggests that the assessor was not simply repeating previous scoring patterns, but was sensitive to the possibility of change across visits. That the QBA dimensions found here were extremely similar to those identified from video by experienced veterinary and farm assurance inspectors, also supports their relevance (Phythian *et al*., 2013a).

Another concern is that qualitative judgments are sensitive to environmental context, which can both be a strength and a potential source of bias (Wemelsfelder *et al.*, 2009; Tuyttens *et al*., 2014; Fleming *et al*., 2015). Random variation in on-farm conditions may have affected QBA scoring levels, and it is difficult to distinguish this from meaningful variation. To counter such effects, it is important to use more than one assessor, and investigate the inter-observer reliability of on-farm QBA assessments (e.g. Andreasen *et al*., 2013). However this is costly and may not be feasible, and it is thus advisable to always apply QBA in combination with other health and welfare measures. The potential for inadvertent observer bias is not exclusive to qualitative methods, but applies equally to quantitative measures (Tuyttens *et al*., 2014). To increase the robustness of on-farm welfare assessments it is crucial to use trained pools of assessors, who are experienced in assessing sheep over the seasons of the year, across a range of production systems, and are able to distinguish meaningful from random fluctuations in sheep expression. If further research were to uphold the positive results found in this study, QBA could potentially be applied as a day-to-day management tool on sheep farms, and be used to communicate welfare values to farmers, shepherds, and consumers.

4.5 Conclusion

The results of this study generally indicate that QBA was capable of identifying expressive dimensions that distinguished meaningfully between sheep demeanour within- and between farms and across the seasons of the year, and correlated significantly with important physical indicators of sheep health. A strong negative correlation was found between PC1 ‘mood’ scores and levels of lameness and ‘dull physical demeanour’, indicating that the latter clinical signs of compromised health also had a wider deleterious effect on the sheep’s emotional state.

Acknowledgments

This study was conducted as part of the Defra funded AW1025 study (Development of indicators for the on-farm assessment of sheep welfare). The authors gratefully acknowledge all the farmers involved for their time and effort in participating with this study.

References

Abramson, Z.H., Abramson, J.H. 2008. Validity. In: Research methods in community medicine: surveys, epidemiological research, programme evaluation, clinical trials. Sixth edition. John Wiley and Sons Ltd, Chichester, England, pp 161-178.

Andreasen, S.N., Wemelsfelder, F., Sandøe, P., 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**.**

Bassler, A.W., Arnould, C. Butterworth, A., Colin, L., De Jong, I.C., Ferrante, V., Ferrari, P., Haslam, S., Wemelsfelder, F., Blokhuis, H.J. 2013. Potential risk factors associated with contact dermatitis, lameness, negative emotional state, and fear of humans in broiler chicken flocks. Poultry Science 92, 2811-2826.

Bokkers, E.A.M., de Vries, M., Antonissen, I.C.M.A., de Boer, I.J.M. 2012. Inter- and intra-observer reliability of experienced and inexperienced observers for the Qualitative Behaviour Assessment in dairy cattle. Animal Welfare 21, 307-318.

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

Boivin, X., Tournadre, H., Le Neindre, P. 2000. Hand-feeding and gentling influence early-weaned lambs' attachment responses to their stockperson. Journal of Animal Science 78, 879-884.

Boivin, X., Lensink, J., Tallet, C., Veissier, I. 2003. Stockmanship and farm animal welfare. Animal Welfare 12, 479-492.

Dwyer, C.M. 2008. Genetic and physiological determinants of maternal behavior and lamb survival: Implications for low-input sheep management. Journal of Animal Science 86**,** 246-258.

Fitzpatrick, J., Scott, M., Nolan, A. 2006. Assessment of pain and welfare in sheep. Small Ruminant Research 62, 55-61.

Fleming, P.A., Wickham, S.L., Stockman, C.A., Verbeek, E., Matthews, L., Wemelsfelder, F. 2015. The sensitivity of QBA assessments of sheep behavioural expression to variations in visual or verbal information provided to observers. Animal 9, 878–887.

French, N.P., Wall, R., Morgan, K.L. 1994. Lamb tail docking: a controlled field study of the effects of tail amputation on health and productivity. Veterinary Record 134, 463-467.

Greenwald, A.G., Pratkanis, A.R., Leippe, M.R., Baumgardner, M.H. 1986. Under what conditions does theory obstruct research progress. Psychological Bulletin 93, 216-299.

Joliffe, I.T. 2002. Principal component analysis. Second edition. Springer series in statistics. Springer-Verlag, New York.

King, E.M., Green, L.E. 2011. Assessment of farmer recognition and reporting of lameness in adults on 35 lowland sheep flocks in England. Animal Welfare 20, 321-328.

Le Neindre, P., Bovin, X., Boissy, A. 1996. Handling of extensively kept animals. Applied Animal Behaviour Science 49, 73-81.

Long, J.S., Freese, L. 2006. Regression models for categorical dependent variables. Using Stata. Stata Press Publications, Texas, USA.

Mellor, D.J. 2012. Animal emotions, behaviour and the promotion of positive welfare states. New Zealand Veterinary Journal 60, 1-8.

Mendl, M., Burman, O.H.P., Paul, E.S. 2010. An integrative and functional framework for the study of animal emotion and mood. Proceedings of the Royal Society B: Biological Sciences 277, 2895–2904.

Minero, M., Tosi, M.V., Canali, E., 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.

Napolitano, F., Knierim, U., Grasso, F., De Rosa, G. 2009. Positive indicators of cattle welfare and their applicability to on-farm protocols. Italian Journal of Animal Science 8, 355-365.

Phythian, C.J., Cripps, P.J., Michalopoulou, E., Jones, P.H., Grove-White, D., Clarkson, M.J., Winter, A.C., Stubbings, L.A., Duncan, J.S. 2012. Reliability of indicators of sheep welfare assessed by a group observation method. The Veterinary Journal 193, 257–263.

Phythian, C., Michalopoulou, E., Duncan, J., Wemelsfelder, F. 2013a. Inter-observer reliability of Qualitative Behavioural Assessments of sheep. Applied Animal Behaviour Science 144, 73–79.

Phythian, C.J., Cripps, P.J., Grove-White, D., Jones, P.H., Michalopoulou, E., Duncan, J.S. 2013b. Observing lame sheep: evaluating test agreement between group-level and individual animal methods of assessment. Animal Welfare 22, 417-422.

Raadsman H.W., Egerton, J.R. 2013. A review of footrot in sheep: Aetiology, risk factors and control methods. Livestock Science 156, 106-114.

Rousing, T., Wemelsfelder, F. 2006. Qualitative assessment of social behaviour of dairy cows housed in loose housing systems. Applied Animal Behaviour Science 101, 40-53.

Rutherford K.M., Donald R.D., Lawrence A.B., Wemelsfelder F. 2012. Qualitative Behavioural Assessment of emotionality in pigs. Applied Animal Behaviour Science 139, 218-224.

Stockman, C.A., Collins, T**.,** Barnes, A.L**.,** Miller, D**.,** Wickham, S.L**.,** Beatty., D.T**.,** Blache, D., Wemelsfelder, F., Fleming, P.A**.,** 2011. Qualitative behavioural assessment and quantitative physiological measurement of cattle naive and habituated to road transport. Animal Production Science 51, 240-249.

Temple, D., Manteca, A., Verlade, A., Dalmay, A. 2011. Assessment of animal welfare through behavioural parameters in Iberian pigs in intensive and extensive conditions. Applied Animal Behaviour Science 131, 29-39.

Temple, D., Manteca, X., Dalmau, A., Velarde, A. 2013. Assessment of test–retest reliability of animal-based measures on growing pig farms. Livestock Production Science 151, 35-45.

Tuyttens, F.A.M., de Graaf S., Heerkens, J.L.T., Jacobs, L., Nalon, E., Ott, S., Stadig, L., Van Laer, E., Ampea, B. 2014. Observer bias in animal behaviour research: can we believe what we score, if we score what we believe? Animal Behaviour 90, 273-280.

Wemelsfelder, F., Lawrence, A.B. 2001. Qualitative assessment of animal behaviour as an on-farm welfare-monitoring tool. Acta Agriculturae Scandinavica Section A-Animal Science 30, 21-25.

Wemelsfelder, F., Farish, M. 2004. Qualitative categories for the interpretation of sheep welfare: a review. Animal Welfare 13, 261-268.

Wemelsfelder, F., Nevison, I., Lawrence, A.B. 2009. The effect of perceived environmental background on qualitative assessments of pig behaviour. Animal Behaviour 78, 477-484.

Wemelsfelder, F., Mullan, S. 2014. Applying ethological and health indicators to practical animal welfare assessment. In: Animal Welfare: focusing on the future (Ed. D.J. Mellor), OIE Scientific and Technical Review, 33, 111-120.

Wickham, S., Collins, T., Barnes, A.L., Miller, D.W., Beatty, D.T., Stockman, C., Blache, D., Wemelsfelder, F. 2012. Qualitative behavioural assessment of transport-naïve and transport habituated sheep. Journal of Animal Science 90, 4523-4535.

Wiseman-Orr, M.L., Scott, E.M., Reid, J., Nolan, A.M. 2006. Validation of a structured questionnaire as an instrument to measure chronic pain in dogs on the basis of effects on health-related quality of life. American Journal of Veterinary Research 67, 1826-1836.

Figure Captions

Figure 1: Loadings of the 12 QBA terms on PC1 and PC2 of the PCA analysis. Axes reflect arbitrary scaling values.

Figure 2: Loadings of the combined PCA analysis of 12 QBA terms and 7 physical health indicators on PC1 and PC2. Axes reflect arbitrary scaling values.

Figure 3: Distributions of QBA PC1 and PC2 scores, and of lameness and breech soiling percentages, for the 12 sheep farms across the 6 visits. Visit periods were: 1. May-June 2009, 2. July-August 2009, 3. September-October 2009, 4. November-December 2009, 5. January-February 2010, and 6. March-April 2010.

Table 1: Overview of assessment visit periods

Visit	Study period	Season	Production stage
1	May – June 2009	Spring/Summer	Post-lambing
2	July – August 2009	Summer	Weaning
3	September – October 2009	Autumn	Mating
4	November – December 2009	Autumn/Winter	Early pregnancy
5	January – February 2010	Winter	Mid-pregnancy
6	March – April 2010	Spring	Lambing

Table 2: Study farm details

Farm ID	Farm type	Flock size	Farm purpose	Farm assurance scheme member	Farming system
1	Lowland	850	Commercial	No	Conventional
2	Lowland	260	Commercial	No	Conventional
3	Hill	24	Small-holding	No	Conventional
4	Lowland	250	Commercial	No	Conventional
5	Lowland	210	Commercial	Yes	Conventional
6	Lowland	280	Commercial	No	Conventional
7	Lowland	600	Commercial	No	Conventional
8	Upland	450	Commercial	Yes	Organic
9	Hill	320	Commercial	Yes	Conventional
10	Hill	800	Commercial	Yes	Conventional
11	Hill	1100	Commercial	Yes	Conventional
12	Hill	1260	Commercial	Yes	Conventional

Table 3: Prevalence of observed physical health and welfare indicators in 12 sheep flocks over 6 visits.

Indicator	**Mean proportion of sheep affected (%) at each of the six assessment visits (95% CI)**						**Mean proportion** (%) **affected across study**
	1 May-June 2009	2 July-August 2009	3 Sept-Oct 2009	4 Nov-Dec 2009	5 Jan-Feb2010	6 March-April 2010	
Dull demeanour	0.17 (-0.07 – 0.41)	0.71 (0.02 – 1. 40)	1.62 (0.01 – 3.21)	0.64 (-0.78 – 2.06)	3.10 (0.04 – 6.17)	3.70 (-0.65 – 8.05)	1.52 (0.77 – 2.77)
Coughing	0	0.18 (-0.08 – 0.43)	0	0.13 (-0.16 – 0.41)	0.28 (-0.11 – 0.67)	0.11 (-0.12 – 0.33)	0.87 (0 – 2.03)
Pruritis	0	0.15 (-0.10 – 0.39)	0.05 (-0.07 – 0.17)	0	0	0.29 (-0.18 – 0.77)	0.12 (0.01 – 0.24)
Wool loss	1.86 (-0.65 – 4.37)	0.18 (-0.09 – 0.44)	0.40 (-0.07 – 0.88)	0	0.85 (0.13 – 1.56)	2.64 (-0.12 – 5.41)	0.74 (0.29 – 1.18)
Lameness	7.23 (3.68 – 10.78)	14.59 (8.79– 20.41)	16.77 (5.62– 27.92)	12.51 (7.14 – 17.89)	17.86 (7.83 – 27.90)	11.29 (4.45 – 18.11)	13.39 (10.53 – 16.22)
Breech soiling	23.83 (11.86 – 35.81)	12.42 (8.69– 16.15)	7.91 (3.25– 12.56)	13.92 (3.08– 24.76)	7.76 (1.28– 14.24)	14.63 (7.97 – 21.27)	13.41 (10.29 – 16.53)
Abdominal soiling	9.24 (-4.37– 22.86)	0	0.11 (-0.14 – 0.37)	20.17 (-0.74 – 41.09)	0.90 (-0.60 – 2.40)	0.37 (-0.50 – 1.20	5.13 (1.09 – 9.18)

Table 4: Regression parameters describing the effect of visit period (1 to 6) on PC1 and PC2 flock scores, for both the QBA and combined QBA/physical indicator analyses, and for physical indicators of lameness and breech and abdominal soiling.

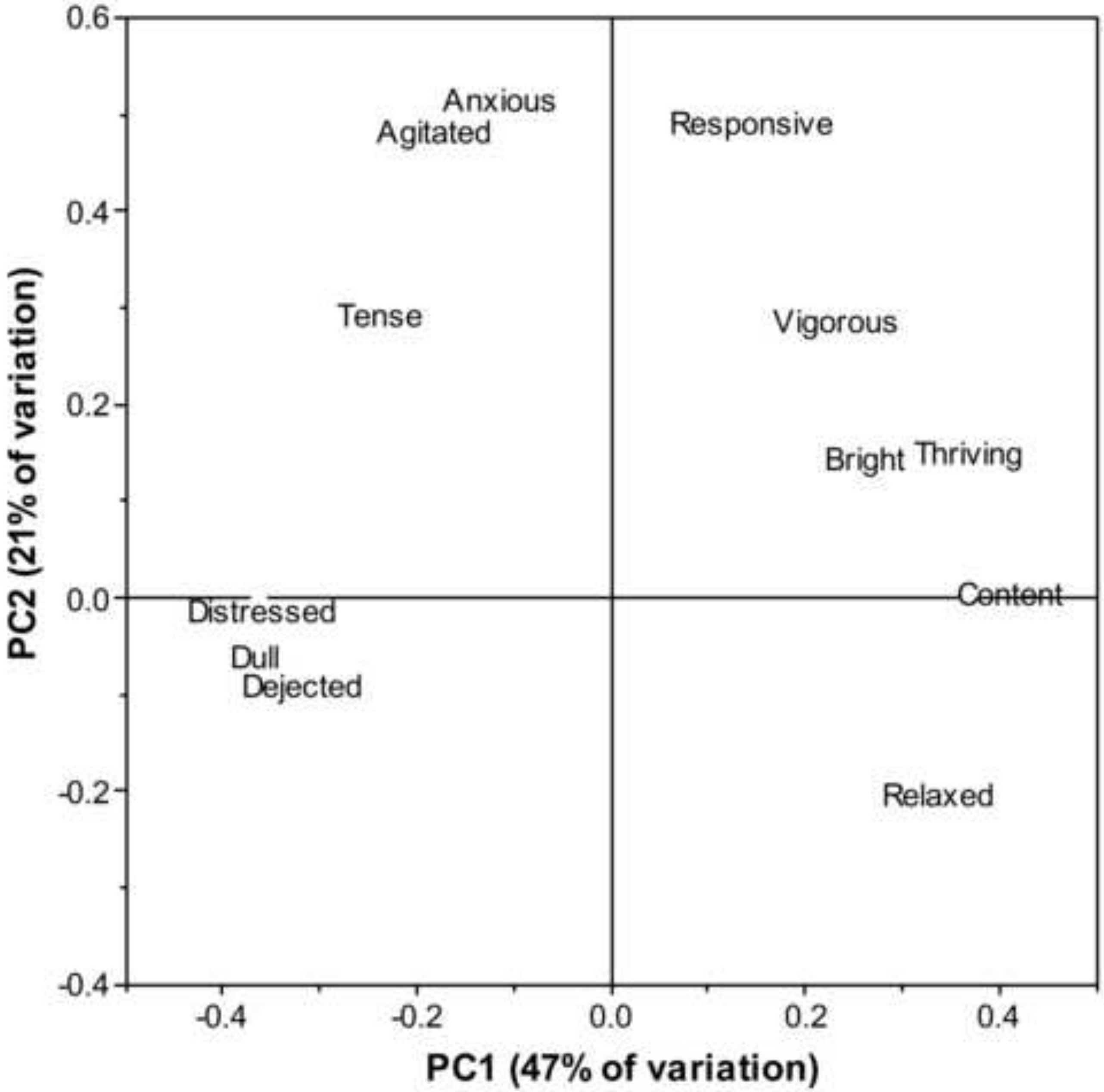
Principal Component	Study visit	β	95% CI	p-value
QBA PC1[a]	**For all 6 visits**	**-**	**-**	**p=0.155**
	1	0.21	-1.11 – 1.54	-
	2	-0.55	-1.40 – 0.29	0.197
QBA PC1	3	-0.54	-1.47 – 0.39	0.255
	4	0.11	-0.84 – 1.06	0.817
	5	-0.56	-1.51 – 0.40	0.251
	6	0.27	-0.68 – 1.23	0.576
QBA PC2	**For all 6 visits**	**-**	**-**	**p<0.001**
	1	-1.25	-2.02 – -0.49	-
	2	1.82	1.14 – 2.49	<0.001
QBA PC2	3	2.04	1.19 – 2.89	<0.001
	4	1.93	0.99 – 2.87	<0.001
	5	1.01	0.02 – 1.99	0.046
	6	0.73	-0.29 – 1.74	0.160
Combined PC1[b]	**For all 6 visits**	**-**	**-**	**p=0.1982**
	1	-0.33	-1.82 – 1.16	-
	2	0.52	-0.45 – 1.51	0.292
Combined QBA/ physical indicators PC1	3	0.60	-0.49 – 1.69	0.277
	4	-0.23	-1.34 – 0.86	0.679
	5	0.87	-0.24 – 1.99	0.125
	6	0.20	-0.91 – 1.32	0.723
Combined PC2	**For all 6 visits**	**-**	**-**	**p<0.001**
	1	-1.46	1.36 – 2.79	-
Combined QBA/physical indicators PC2	2	2.07	1.39 – 3.13	<0.001
	3	2.26	1.39 – 3.27	<0.001
	4	2.33	0.46 – - 2.39	<0.001

	5	1.42	-0.31– 1.65	0.004
	6	0.67	-0.31– 1.65	0.178
	For all 6 visits	**-**	**-**	**p=0.0249**
Lameness	1	7.23	0.72– 13.74	-
	2	7.37	0.49– 14.23	0.036
	3	9.53	2.90– 16.17	0.005
	4	5.28	-1.37 – 11.93	0.120
	5	10.63	3.98– 17.28	0.002
	6	4.05	-2.59– 10.70	0.232
	For all 6 visits	**-**	**-**	**p=0.0008**
Breech soiling	1	23.83	16.92– -4.16	-
	2	-11.40	-18.65– -4.17	0.002
	3	-15.92	-24.21– -7.64	0.001
	4	-9.90	-18.49– -1.33	0.024
	5	-16.06	-24.74– -7.40	0.001
	6	-9.21	-17.90– -.51	0.038
	For all 6 visits	**-**	**-**	**p=0.0048**
Abdominal soiling	1	9.24	0.53 – 17.95	-
	2	-9.24	-21.64– 3.15	0.144
	3	-9.12	-21.44– 3.17	0.146
	4	10.93	-1.39– 23.25	0.082
	5	-8.34	-20.66 – 3.97	0.184
	6	-8.87	-21.19– 3.45	0.158

[a] Principal Component obtained through PCA of QBA scores only

[b] Principal Component obtained through PCA of QBA and physical indicator scores

Figure 1



0.6
0.4
0.2
0.0
-0.2
-0.4
-0.4
-0.2
0.0
0.2
0.4
PC2 (21% of variation)
PC1 (47% of variation)
Anxious
Agitated
Responsive
Tense
Vigorous
Bright
Thriving
Content
Distressed
Dull
Dejected
Relaxed

Figure 2

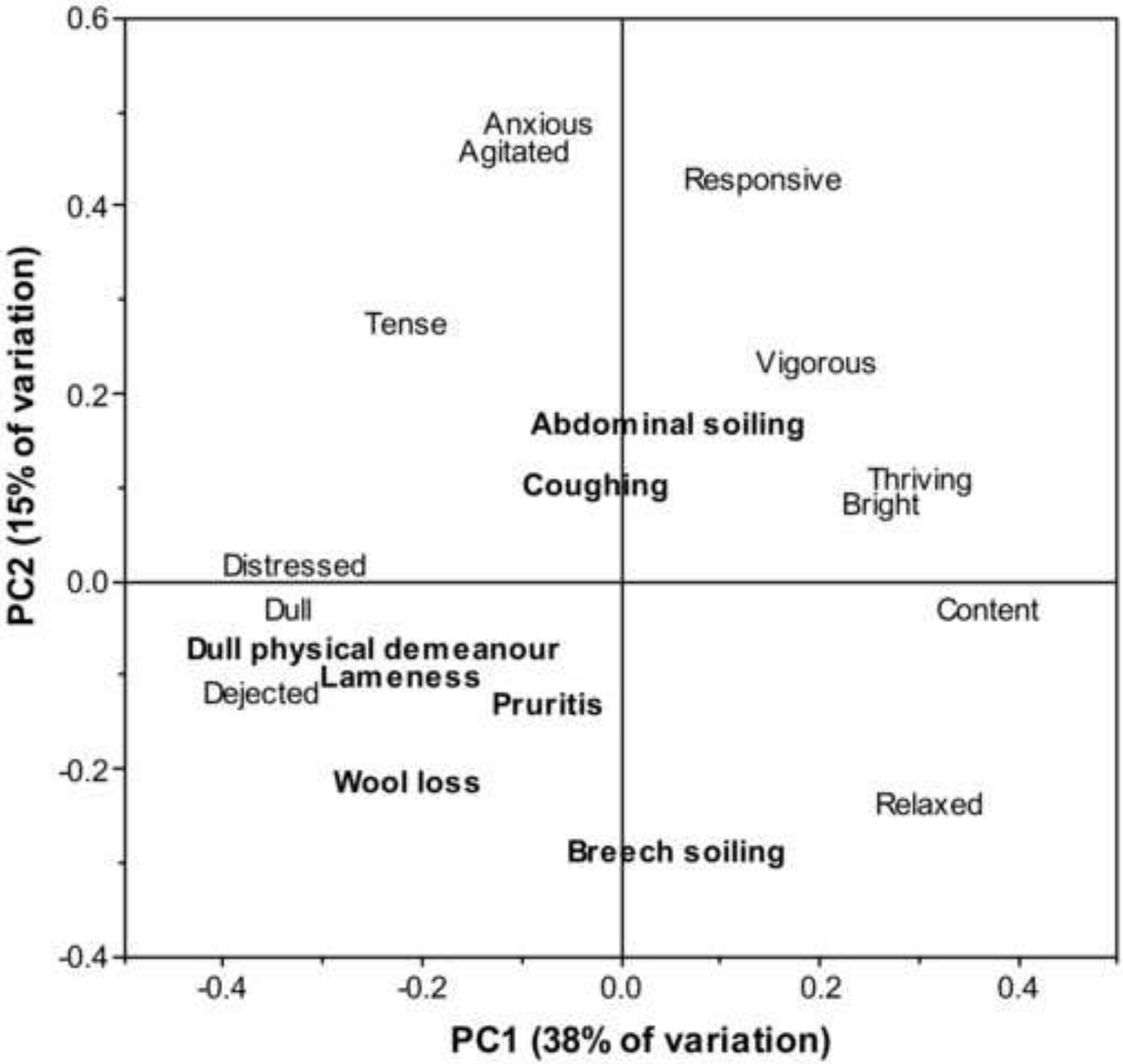
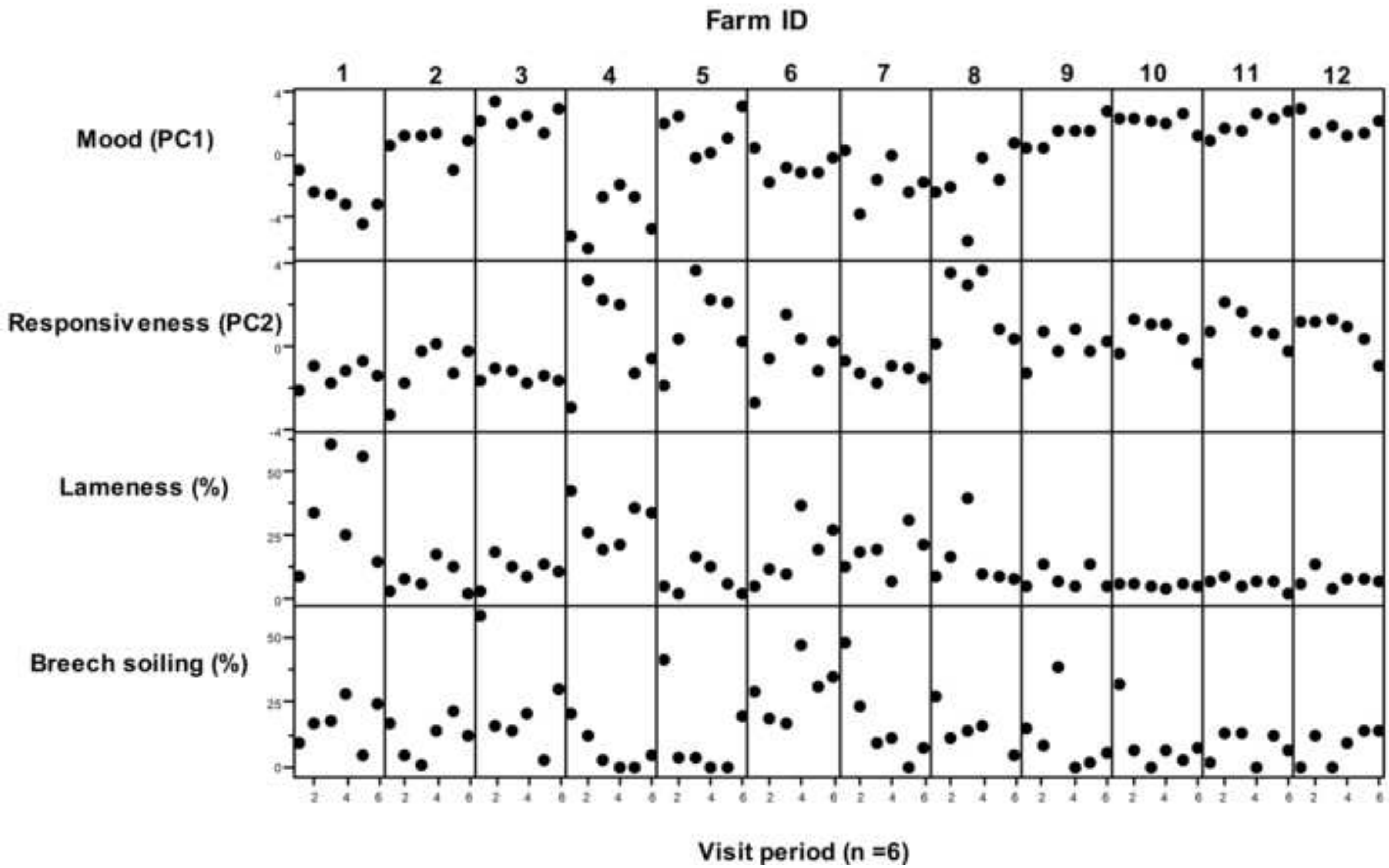

Anxious
Agitated
Responsive
Tense
Vigorous
Abdominal soiling
Coughing
Thriving
Bright
Distressed
Dull
Content
Dull physical demeanour
Lameness
Dejected
Pruritis
Wool loss
Relaxed
Breech soiling
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.4
-0.2
0.0
0.2
0.4
PC2 (15% of variation)
PC1 (38% of variation)

Figure 3


Farm ID
1
2
3
4
5
6
7
8
9
10
11
12
Mood (PC1)
Responsiveness (PC2)
Lameness (%)
Breech soiling (%)
Visit period (n =6)