

A pilot investigation of Farm Assurance assessors' attitude to farm animal welfare as a confounding factor to training in pig welfare outcome measures

S Mullan^{*†}, SA Edwards[‡], A Butterworth[†], HR Whay[†] and DCJ Main[†]

[†] Department of Clinical Veterinary Science, University of Bristol Veterinary School, Langford, Somerset BS40 5DU, UK

[‡] Newcastle University, School of Agriculture Food & Rural Development, Newcastle-upon-Tyne NE1 7RU, UK

* Contact for correspondence and requests for reprints: Siobhan.mullan@bris.ac.uk

Abstract

The effect of Farm Assurance (FA) assessors' attitude to farm animal welfare on the inter-observer reliability of some welfare outcome measures achieved following training was investigated as part of a larger project examining the feasibility and benefits of the incorporation of some on-farm welfare outcome assessments into UK Pig Farm Assurance Schemes. A total of thirty-one FA assessors were trained in three training sessions to assess the following welfare outcome measures: body lesions, tail lesions, severe tail lesions, lameness and pigs requiring hospitalisation. Assessment of photographs, live observations of individual pigs and pens of pigs were used to generate inter-observer reliability data. A previously validated farm animal welfare questionnaire was used to assess the FA assessors' attitudes to farm animal welfare. Principal Component Analysis of FA assessor scores for this questionnaire resulted in two major components, with component 1 termed 'pigs have mental welfare' and component 2 termed 'people-centric, pigs as profit'. FA assessors demonstrated a range in attitudes to farm animal welfare and, when assessing the same pigs, recorded a range in prevalence of welfare outcome measures and degree of agreement with a gold standard following training. There were only seven out of a possible 98 significant correlations between the FA assessor scores for components 1 and 2 and their recorded prevalence of welfare measures and levels of agreement with a gold standard. In particular, FA assessors' scores for component 1 were significantly positively correlated with the recorded prevalence for pigs requiring hospitalisation in two of the three training sessions although there was no effect on the agreement with a gold standard for this measure. These results indicate that training in welfare outcomes, defined by a standard protocol, is relatively unconfounded by observer attitudes to farm animal welfare. To obtain better levels of agreement between assessors, and therefore more reliable data, it is recommended that FA schemes concentrate their resources on providing good quality training in a well-defined protocol and reliability testing and that they do not need to attempt to account for the attitudes of the FA assessors to farm animal welfare.

Keywords: animal welfare, attitude to animals, Farm Assurance, inter-observer reliability, pig, welfare outcome

Introduction

Farm Assurance (FA) aims to provide assurances to the market over food safety, animal welfare and environmental concerns and was initiated by the UK pig industry more than a decade ago (Main & Green 2000). Farm Assurance standards for assessing welfare have, up to now, been mainly based on an assessment of the resources available to the pigs, such as the quality of the physical environment and stockmanship (welfare inputs). It is recognised that assessing welfare outcomes, such as direct observations of an animal's behaviour and physical condition, may be better able to give an insight into how the animal experiences the effect of the inputs (Main *et al* 2007). For example, the Farm Animal Welfare Council (FAWC) report (FAWC 2005) recommends that Farm Assurance schemes:

“work towards refining their standards and inspection procedures to achieve an increasing inclusion of welfare outcomes, so as to provide both a better reflection of

the welfare of the animals within a production system and the level of stockmanship on the farm”.

This study was part of a larger project examining the feasibility and benefits of incorporating some on-farm welfare outcome assessments into the harmonised UK Pig FA Schemes, run by Assured British Pigs (ABP) and Genesis Quality Assurance (GQA) (ABP 2007; GQA 2007). Trained FA assessors audit these standards on farms, spending part of their time observing a sample of the pigs. In order to record reliable data for the whole farm when formally including welfare outcomes into FA schemes, it is important that the sample of pigs observed is sufficiently representative of the farm as a whole (see Mullan *et al* 2009), that the assessment is not affected by factors such as the time of day (see Mullan *et al* 2011) and that there is consistency of assessment between FA assessors (Keeling 2007).

Previous studies have documented the inter-observer agreement following training in welfare outcome measures (eg Petersen *et al* 2004; Kristensen *et al* 2006). In one study, the effect of the amount of training on the inter-observer agreement achieved was reported, with more training sessions producing better levels of agreement between observers assessing dairy cow lameness (March *et al* 2007). This study aimed to investigate whether an FA assessor's attitude to animal welfare would confound the training in welfare outcome measures, and therefore the inter-observer agreement achieved.

Materials and methods

Three training and testing sessions for FA assessors were organised at pig farms in different regions of the UK. There were a total of five trainers used in the sessions. Sessions one and two were attended by the same four trainers and for session three there were three trainers involved, one of whom had not taken part in the previous two sessions. A total of 31 FA assessors attended the training sessions with 13, 9 and 9 assessors at sessions one, two and three, respectively.

Following a brief introduction to the wider project, the FA assessors completed a modified Edinburgh Farm Animal Welfare Scale (EFAWS) questionnaire to determine peoples' attitudes to farm animal welfare, developed and validated at Edinburgh University, UK, using pig farmers as part of the survey population (Austin *et al* 2005). The following statements to be rated for agreement were removed from the previously published questionnaire as they were particularly specific to farmers: 'I often visit my pigs just to see how they are getting on'; 'I like to keep my staff informed about changes to the law on animal welfare'; 'My staff and I agree on standards of care, I do not consider costs before calling the vet to a sick animal'; 'Not being able to transport lame animals limits my farming activity'; 'I try to put the more patient/caring staff where stock are giving birth'; 'I consider individuals within my herd to be like pets'. All assessors were encouraged to rate their agreement, on a five-point scale, with the statements outlined in Table 1 truthfully. They then rated various welfare issues for pigs on a four-point scale of severity of the problem to the pig (where 1 = no problem and 4 = great problem). The issues included tail biting for the bitten pig and the biter, lameness, dirtiness, poor growth rate, fighting, boredom, lack of a manipulable material but presence of a manipulable object, lack of a manipulable material and object, body lesions, uncomfortable lying area, bursae, fear of people including stockmen, non-removal to a hospital pen of sick or injured pigs.

The welfare outcome parameters that the FA assessors were trained in were: body lesions (one side of a pig assessed for presence or absence of a ≥ 3 cm linear skin wound or ≥ 1 cm diameter circular wound), tail lesions (presence or absence of a skin wound on any part of the tail), severe tail lesions (at least the tail hanging down with fresh blood or scabs covering the tip), lameness (pigs with lameness score 3 or more — any pig that 'when standing has at least an uneven posture, will not bear full weight on the affected limb and appears to be standing on its toes; when moving there is a

shortened stride with minimum weight-bearing on the affected limb but will still trot and gallop' [Main *et al* 2000]) and pigs requiring hospitalisation (any sick or injured pigs that would benefit from being 'temporarily isolated in suitable accommodation with, where appropriate, dry comfortable bedding' (The Welfare of Farmed Animals [England] Regulations [The Welfare of Farmed Animals (England) Regulations 2007])). The training was assisted by a pictorial and written guide to the parameter thresholds and started with a classroom familiarisation of the guide and recording sheets lasting approximately 30 mins. This was followed by on-farm discussions of individual pigs in groups of up to four assessors per trainer. Finally, training in the use of hand-held tally counters to count the numbers of affected pigs, from inside each pen, included a practice on at least five pens. The total training time was approximately 2 h.

The reliability assessments consisted of three parts undertaken by both the gold standard trainer (SM) and all the assessors. There was an assessment of approximately 50 photographs for both the assessment of tail lesions and body lesions. Approximately 20 individual pigs were directly assessed for body lesions (left side only), tail lesions, severe tail lesions, lameness and pigs requiring hospitalisation. In addition, the pen prevalences of body lesions, tail lesions, severe tail lesions, lameness and pigs requiring hospitalisation for approximately 20 pens of pigs were recorded. All training and reliability assessments took place on finishing pigs in pens of 25 pigs or less. The pigs in session one were housed in fully slatted accommodation and fed a liquid diet, those in session two were housed on straw with dry food and those in session three were housed on a solid floor without a manipulable material and fed a liquid diet.

Statistical analysis

Statistical analysis was performed using SPSS vs 14.0. The agreements between assessors and the gold standard trainer for both the individual pigs and the photographs were tested using Cohen's Kappa. The pen prevalences were tested for agreement with the gold standard trainer using Kendall's *W* coefficient of concordance. The EFAWS was analysed using principal component analysis (PCA, covariance matrix without rotation). Principal Component Analysis transforms data consisting of a range of variables, which may correlate, into a smaller number of usually uncorrelating variables, termed 'components' (Shaw 2003). Spearman's *rho* was used to identify correlations between the component scores produced and the Kendall's *W* or Cohen's kappa value for agreement of the assessors with the gold standard trainer, and the actual prevalences of welfare outcome measures recorded by assessors. It was also used to determine any correlations between the component scores and the difference in recorded prevalence for the same pigs between the assessor and the gold standard trainer, as a proportion of the gold standard prevalence (termed 'prevalence residual'). This prevalence residual is a measure of how close the FA assessor recorded prevalence is to the gold standard trainer's recorded prevalence for the same pigs, regardless of overall statistical agreement. For example, if the gold standard trainer's recorded prevalence was 10% and one FA

Table 1 Loadings of the items in the animal welfare questionnaire on 5 components.

| | 1 | 2 | 3 | 4 | 5 |
|---|---------------|--------------|---------------|--------------|--------|
| It is important for an animal's psychological needs to be met | 0.735 | 0.299 | -0.037 | 0.052 | -0.145 |
| I encourage discussion of animal welfare issues with farm staff | 0.641 | -0.001 | 0.21 | -0.091 | -0.015 |
| For the animals' sake it is important to minimise the distance they are transported to the market/abattoir | 0.602 | 0.343 | 0.318 | 0.065 | 0.246 |
| I would not want to be without household pets | 0.597 | 0.558 | 0.053 | 0.11 | 0.357 |
| It is important to check pigs last thing at night even when there are no health problems | 0.572 | 0.507 | 0.164 | -0.05 | -0.27 |
| I prefer not to kill crows/seagulls but do so if strictly necessary | 0.524 | -0.124 | -0.23 | -0.08 | -0.455 |
| A lame animal should always be treated or culled | 0.513 | 0.083 | 0.151 | -0.14 | -0.145 |
| Incurable animals should be put down by a trained person | 0.463 | 0.198 | 0.199 | 0.437 | 0.25 |
| Farm animals should be kept in as natural an environment as possible | 0.462 | -0.473 | -0.115 | -0.121 | 0.027 |
| I can tell animals in the breeding herd apart by their natural characteristics | 0.436 | 0.505 | -0.154 | -0.183 | -0.149 |
| It is important for animals to be able to perform their natural behaviour | 0.434 | -0.37 | 0.275 | -0.215 | 0.173 |
| The idea of a 'natural environment' applies to farm animals as well as wild animals | 0.404 | -0.324 | -0.302 | 0.413 | 0.121 |
| Companion animals (pets) enhance our quality of life | 0.371 | -0.017 | 0.37 | 0.039 | 0.411 |
| I do not like to kill foxes and actively avoid doing so | 0.367 | 0.057 | -0.575 | 0.431 | -0.285 |
| I give names to some of the animals in the breeding herd | 0.336 | 0.198 | -0.244 | -0.29 | 0.567 |
| I do not like to kill crows/seagulls and actively avoid doing so | 0.280 | -0.016 | -0.108 | 0.752 | -0.337 |
| My own standards of animal welfare are more important to me than the views of others | 0.254 | 0.119 | -0.087 | 0.492 | 0.525 |
| Using traps to control pests (other than mice) is cruel/unacceptable | 0.251 | -0.082 | -0.397 | 0.596 | -0.083 |
| I like to be informed about new knowledge relating to animal welfare | 0.233 | 0.718 | 0.065 | -0.181 | -0.027 |
| It is important for children to have the experience of keeping pets | 0.230 | -0.003 | 0.246 | -0.145 | 0.364 |
| A pet should be treated as a member of the family | 0.226 | 0.663 | -0.13 | 0.187 | 0.248 |
| Rats should be controlled in England by systematic shooting or trapping | 0.222 | 0.294 | 0.692 | 0.026 | 0.01 |
| It is important to know the individual character of a pig in order to assess whether it is acting out of character and perhaps ill or in pain | 0.211 | 0.048 | 0.198 | 0.675 | -0.163 |
| I like to be informed about changes to the law on animal welfare | 0.197 | 0.511 | -0.047 | -0.473 | 0.086 |
| Government legislation on welfare is helpful to the farmer | 0.189 | 0.078 | 0.431 | -0.06 | -0.04 |
| I consider the possible impression on visitors to the farm when making welfare decisions | 0.184 | 0.045 | 0.044 | 0.016 | -0.323 |
| I would put down suffering incurable animals myself | 0.169 | 0.58 | 0.106 | 0.272 | 0.037 |
| Mice should be controlled in England by systematic shooting or trapping | 0.151 | 0.084 | 0.372 | 0.447 | 0.113 |
| Public opinion should not dictate welfare standards on-farm | 0.098 | 0.027 | 0.204 | -0.307 | 0.03 |
| I prefer not to kill rats but do so if strictly necessary | 0.042 | 0.276 | -0.584 | 0.031 | 0.264 |
| An animal that is physically fit must have good welfare | 0.016 | 0.432 | 0.131 | 0.042 | -0.512 |
| There are differences between staff in their care of animals | 0.005 | 0.148 | -0.437 | 0.107 | 0.22 |
| I prefer not to kill mice but do so if strictly necessary | -0.004 | 0.417 | -0.593 | -0.046 | -0.037 |
| I do not like to kill rabbits but do so if strictly necessary | -0.056 | -0.156 | -0.604 | -0.066 | -0.11 |
| The export of live animals to the continent for food is cruel and should not be permitted | -0.066 | -0.339 | 0.346 | 0.67 | 0.049 |
| Animal welfare organisations are unhelpful to farmers | -0.081 | 0.256 | -0.472 | -0.038 | 0.297 |
| Production efficiency should be the first priority of the farmer | -0.119 | 0.808 | 0.093 | 0.214 | -0.208 |
| Rabbits should be controlled in England by systematic gassing, shooting or trapping | -0.14 | -0.164 | 0.736 | 0.163 | 0.103 |
| Transport of farm animals by road and rail involves little discomfort or pain | -0.16 | 0.391 | 0.06 | 0.068 | -0.566 |
| The Government should pay for improvements in farm animal welfare | -0.164 | 0.133 | 0.489 | -0.138 | 0.147 |
| If an animal is growing well it must be experiencing good welfare | -0.242 | 0.42 | 0.21 | -0.08 | -0.572 |
| Using traps to control mice is more humane than using chemicals | -0.26 | 0.329 | -0.593 | 0.063 | 0.451 |
| An animal that is physically healthy cannot be suffering | -0.286 | 0.252 | 0.023 | 0.355 | 0.214 |
| Public concern about the welfare of animals is greatly exaggerated | -0.294 | 0.031 | 0.068 | -0.235 | -0.088 |
| If an animal is reproducing efficiently its welfare standards must have been good | -0.295 | 0.497 | 0.022 | 0.14 | -0.149 |
| Crows/seagulls should be controlled in England by systematic shooting or trapping | -0.297 | 0.396 | 0.271 | 0.016 | 0.446 |
| It is more important to control disease than to keep farm animals in a natural environment | -0.303 | 0.36 | -0.043 | -0.061 | -0.271 |
| Vets are expensive in relation to the value of commercial stock so it is not worth bothering them, except for problems that affect the whole herd | -0.341 | 0.039 | -0.31 | 0.277 | -0.22 |
| I have no qualms about personally killing rabbits | -0.385 | -0.22 | 0.456 | 0.016 | -0.086 |
| I think of pigs mainly in terms of the profit they will bring in | -0.404 | 0.668 | 0.03 | -0.082 | 0.086 |
| I would rather have a sick pet put down than pay expensive vets' fees | -0.419 | -0.038 | 0.338 | 0.595 | 0.047 |
| I think of my stock mainly in terms of their market value or cost | -0.442 | 0.642 | 0.021 | -0.071 | -0.038 |
| Foxes should be controlled in Scotland by systematic shooting or trapping | -0.529 | 0.262 | 0.4 | 0.024 | 0.135 |
| A sick animal should be left to its own devices as they often recover | -0.649 | 0.068 | -0.114 | 0.096 | 0.344 |
| I tend to think of pigs as being very similar to machines | -0.652 | 0.341 | 0.029 | 0.13 | 0.061 |
| All pigs are the same; dumb animals | -0.795 | -0.122 | -0.137 | 0.209 | 0.094 |

Loadings > 0.6 are presented in bold (n = 25 FA assessors).

Table 2 The agreement of farm assessors with the gold standard trainer for a range of welfare outcome measures and assessment methods.

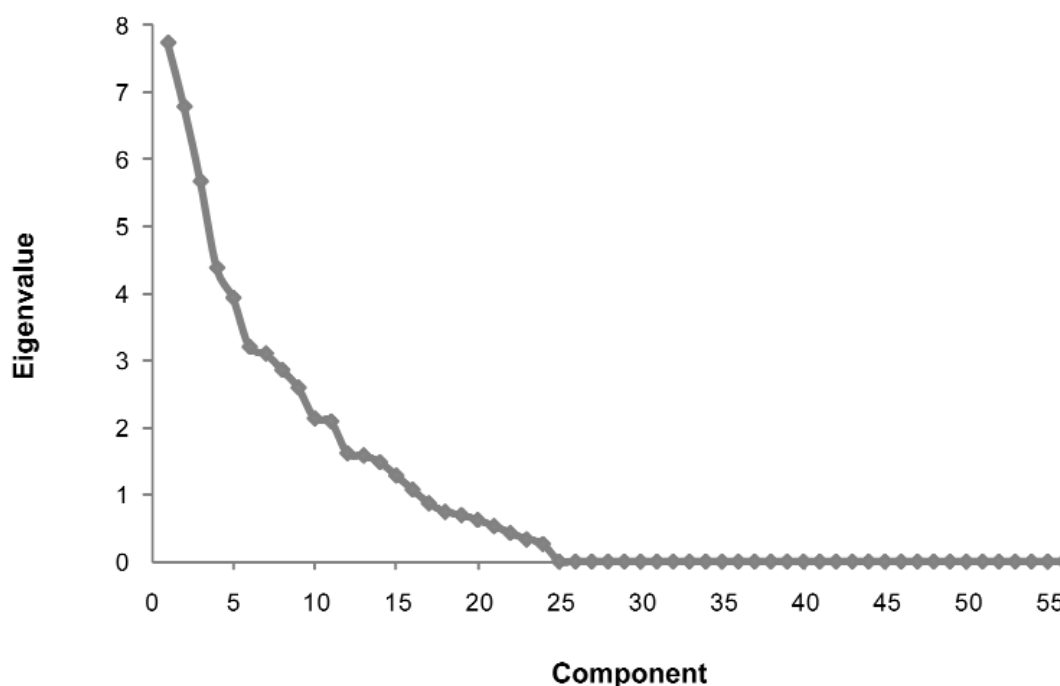
| | Test setting | Agreement statistic used | Number of assessors with the following agreement values with the gold standard trainer | | | | | Total number of assessors |
|--------------------------------|---------------------|--------------------------|--|-----------|-----------|-----------|---------|---------------------------|
| | | | < 0.2 | 0.201–0.4 | 0.401–0.6 | 0.601–0.8 | 0.801–1 | |
| Tail lesions | 50 photos | Cohen's kappa | 1 | 1 | 17 | 11 | 1 | 31 |
| | ~20 pens of pigs | Kendall's W | 0 | 1 | 9 | 15 | 6 | 31 |
| | ~20 individual pigs | Cohen's kappa | 0 | 5 | 6 | 6 | 0 | 17 |
| Severe tail lesions | ~20 pens of pigs | Kendall's W | 0 | 0 | 10 | 10 | 6 | 26 |
| | ~20 individual pigs | Cohen's kappa | 1 | 0 | 0 | 1 | 9 | 11 |
| Body lesions | 45 photos | Cohen's kappa | 2 | 3 | 15 | 8 | 3 | 31 |
| | ~20 pens of pigs | Kendall's W | 0 | 0 | 6 | 19 | 6 | 31 |
| | ~20 individual pigs | Cohen's kappa | 2 | 7 | 5 | 4 | 0 | 18 |
| Lame pigs | ~20 pens of pigs | Kendall's W | 0 | 1 | 9 | 20 | 1 | 31 |
| | ~20 individual pigs | Cohen's kappa | 0 | 1 | 4 | 7 | 6 | 18 |
| Pigs requiring hospitalisation | ~20 pens of pigs | Kendall's W | 0 | 1 | 14 | 12 | 4 | 31 |
| | ~20 individual pigs | Cohen's kappa | 1 | 6 | 3 | 2 | 1 | 13 |

Table 3 Significant Spearman's rho correlations at $P \leq 0.05$ between the two main components derived from the EFAWS and the Farm Assurance assessors' prevalences of welfare measures, agreement with a gold standard and rating of welfare issues.

| | | Component 1: 'Pigs have mental welfare' | | | | Component 2: 'People-centric, pigs as profit' | | | |
|--|--|---|---|--|---------|---|---|--|---------|
| | | Welfare measure | Test statistic used | r_s | P-value | Welfare measure | Test statistic used | r_s | P-value |
| Actual prevalences recorded by FA assessors | Training session 1 | Tail lesions (n = 7) | Overall prevalence of ~20 pens | 0.786 | 0.036 | | | | |
| | | Lameness (n = 7) | Overall prevalence of ~20 pens | 0.882 | 0.008 | | | | |
| | Training session 2 | Pigs requiring hospitalisation (n = 9) | Overall prevalence of ~20 individual pigs | 0.819 | 0.007 | | | | |
| | | Training session 3 | Pigs requiring hospitalisation (n = 9) | Overall prevalence of ~20 individual pigs | 0.718 | 0.029 | Tail lesions (n = 9) | Overall prevalence of ~20 pens | 0.850 |
| FA assessor agreement with gold standard trainer, Cohen's kappa or Kendall's W | Rating of 14 welfare issues (1 = no problem, 4 = great problem) | | Boredom (n = 25) | Score given by FA assessors (1 = no problem, 2 = not much of a problem, 3 = some problem, 4 = great problem) | 0.530 | 0.006 | Fear of people, including stockmen (n = 23) | Score given by FA assessors (1 = no problem, 2 = not much of a problem, 3 = some problem, 4 = great problem) | 0.420 |
| | | Fighting (n = 25) | | 0.430 | 0.032 | | | | |
| | | Bursae (n = 25) | | 0.470 | 0.016 | | | | |
| | Lack of manipulable material but presence of an enrichment object such as a chain (n = 25) | | 0.408 | 0.042 | | | | | |
| | | | | | | Severe tail lesions (n = 8) | Overall prevalence of ~20 pens | -0.769 | 0.025 |
| | | | | | | Lameness (n = 25) | Overall prevalence of ~20 pens, Kendall's W | -0.478 | 0.016 |

There were no significant correlations between components 1 and 2 and Farm Assurance assessors' prevalence residuals with the gold standard trainer.

Figure 1



The scree plot of the components derived from Principal Component Analysis of the result of the modified Edinburgh Farm Animal Welfare Scale (n = 25 assessors).

assessor recorded 5% and another recorded 15% both FA assessors would have a prevalence residual of 50%.

Number of assessors

The results from all assessors are presented where possible in Tables 2 and 3, but sample sizes varied as a result of individual circumstances pertaining at the different training sessions. Six assessors did not complete the EFAWS. All assessors completed the assessment of photographs. In assessing the prevalence of welfare outcomes at a pen level some assessors did not record the prevalence of severe tail lesions. In some sessions, the prevalence of a welfare outcome measure for the 20 individual pigs was recorded by the gold standard trainer as 0 and Cohen's Kappa could not be calculated.

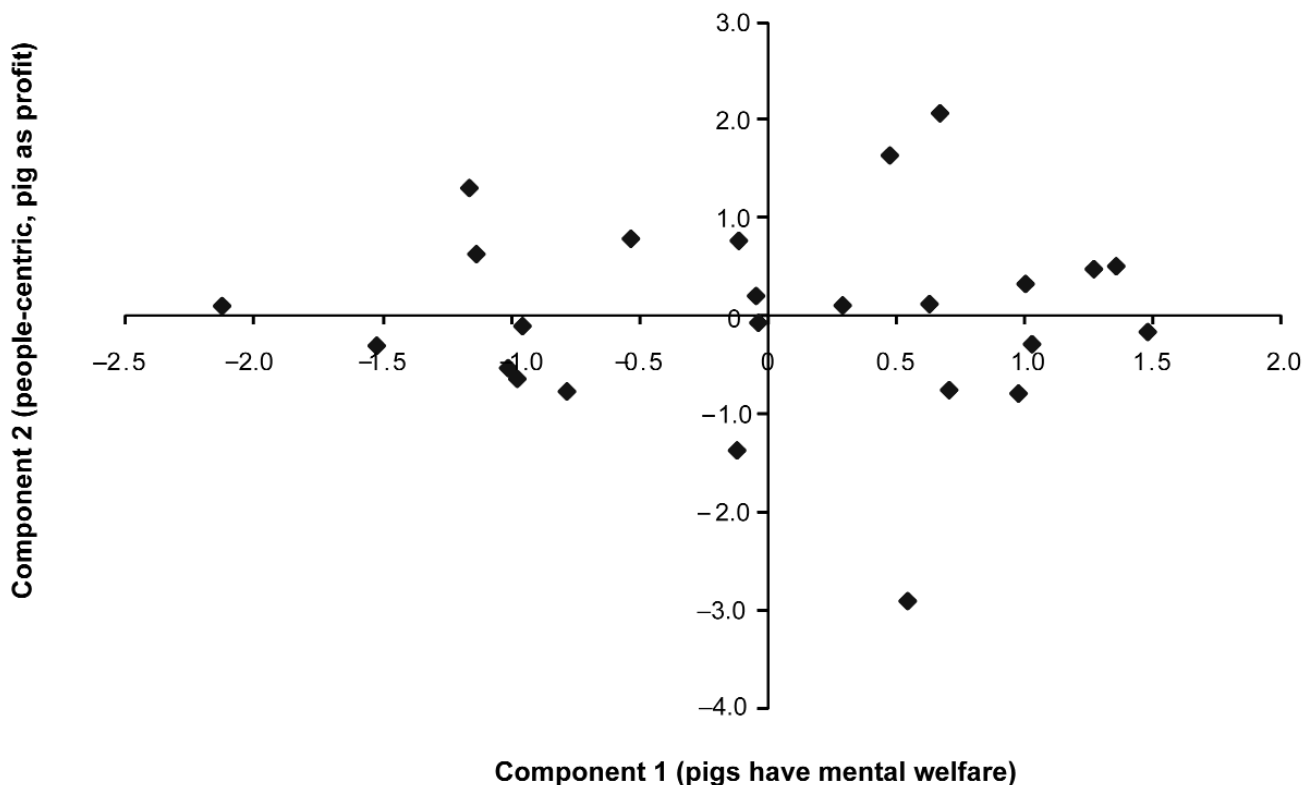
Results

The Edinburgh Farm Animal Welfare Scale (EFAWS)

Twenty-five FA assessors completed the EFAWS questionnaire, seven from session one and nine from sessions two and three. There were 16 components with Eigenvalues greater than one ruling out this criterion for determining the useful components for further analysis. Examination of the Principal Component Analysis scree plot (Figure 1) suggested up to a 5-component solution explaining 50.9% of the variance in the data. The loadings of each question on the 5 components are shown in Table 1. The highest loading items (values over 0.6) on component 1, termed 'pigs have mental welfare' included items relating to the ability of the pig to experience poor welfare (eg 'It is important for an

animal's psychological needs to be met' [positive loading]; 'All pigs are the same — dumb animals' [negative], 'I tend to think of pigs as being very similar to machines' [negative]) but also included items relating to farm management (eg 'I encourage discussion of animal welfare issues with farm staff'). The highest loading items, with values over 0.6, on component 2, termed 'people-centric, pigs as profit', included items relating to the production efficiency of a farm (eg 'Production efficiency should be the first priority of the farmer', 'I think of pigs mainly in terms of the profit they will bring in', 'I think of my stock mainly in terms of their market value or cost' [all positive loading]; 'Farm animals should be kept in as natural environment as possible [negative]) but also included items referring to their knowledge of animal welfare ('I like to be informed about new knowledge relating to animal welfare') and pet-keeping ('A pet should be treated as a member of the family'). Component 3, termed pest control, has its loadings with values over 0.6 from questions about killing pests (eg 'Rabbits should be controlled in England by systematic gassing, shooting or trapping'; 'I prefer not to kill mice but do so if strictly necessary'). Component 4 is a mixed component with no clear theme. The loadings with values over 0.6 are 'I do not like to kill crows/seagulls and actively avoid doing so' and 'It is important to know the individual character of a pig in order to assess whether it is acting out of character and possibly ill or in pain'. Component 5 had no loadings with values more than 0.6. On examination of the 5 components, the 2-component solution was chosen for

Figure 2



The component scores for each Farm Assurance assessor for component 1 (pigs have mental welfare) and component 2 (people-centric, pigs as profit).

further analysis as it contained the components of most relevance to pig welfare assessment.

The PCA scores for each FA assessor for component 1 (pigs have mental welfare) and component 2 (people-centric, pigs as profit) are shown in Figure 2. There was a greater range of scores for component 2 (people-centric, pigs as profit).

Rating of animal welfare issues

The ratings by twenty-five FA assessors of various animal welfare issues are presented in Figure 3. The welfare issues which the highest number of assessors rated as a great problem for a finishing pig were tail biting — for the bitten pig (22 assessors), non-removal to a hospital pen if sick/injured (18 assessors), lameness and body lesions (17 assessors each). The welfare issues that the most assessors rated as less of a problem were dirtiness (8 assessors), tail biting — for the biter (5 assessors) and fear of people, including stockmen (4 assessors).

Agreement between FA assessors and the gold standard trainer

The levels of agreement between FA assessors and the gold standard trainer for a range of welfare outcome measures and test settings are shown in Table 2. The highest agreement levels were found for severe tail lesions of individual pigs (10/11 assessors had a kappa value > 0.6). The

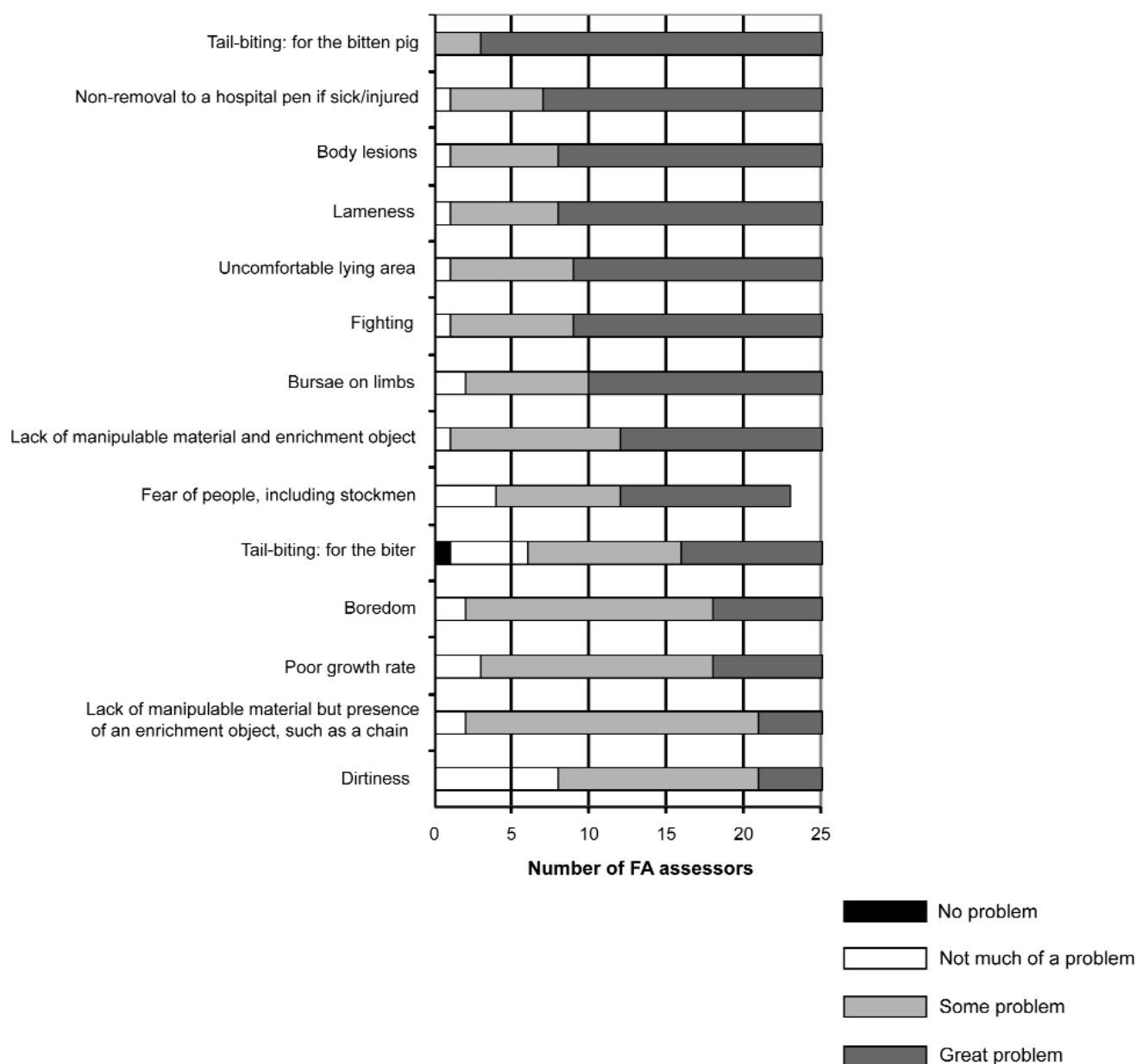
least agreement was shown when examining individual pigs as to whether they required hospitalisation (23% of FA assessors had kappa values > 0.6) or had body lesions (22% of FA assessors had kappa values > 0.6).

The relationship between FA assessor EFAWS component scores and their recorded prevalences of welfare outcome measures and agreement with the gold standard trainer following training

The statistically significant ($P \leq 0.05$) spearman's *rho* correlations between components 1 and 2 and the prevalence of welfare outcome measures recorded by FA assessors and their levels of agreement with the gold standard trainer (Cohen's kappa or Kendall's *W*) are shown in Table 3. Of a total of 49 possible correlations with each component tested, there were four significant correlations with component 1 ('pigs have mental welfare') and three significant correlations with component 2 (people-centric, pigs as profit).

There were positive correlations between FA assessors' values for component 1 (pigs have mental welfare) and the recorded prevalence of pigs requiring hospitalisation when examining ~20 individual pigs in two of the training sessions ($r_s = 0.819, P = 0.007; r_s = 718, P = 0.029$). In one training session there were also significant positive correla-

Figure 3



Ratings of 14 pig welfare issues by farm assurance assessors (n = 25).

tions between component 1 and the recorded prevalence of tail lesions ($r_s = 0.786$, $P = 0.036$) and lame pigs ($r_s = 0.882$, $P = 0.008$) when assessing pens of pigs. However, there were no correlations between levels of agreement or prevalence residuals and component 1.

FA assessors' values for component 2 (people-centric, pigs as profit) were found to correlate significantly and positively with recorded prevalence of tail lesions in pens of pigs in one training session ($r_s = 0.850$, $P = 0.029$), but in that same training session there was a significant negative

correlation with recorded prevalence of pigs with severe tail lesions, effectively a sub-category of tail lesions, ($r_s = -0.769$, $P = 0.025$). Component 2 had the only significant correlation with the level of agreement with the gold standard where FA assessors' values for 'people-centric, pigs as profit' were negatively correlated with their Kendall's W -values for the prevalence of lameness in pens of pigs ($r_s = -0.478$, $P = 0.016$). There were no significant correlations between the FA prevalence residual with the gold standard trainer and component 2.

Correlations between the rating of welfare issues and the EFAWS

The significant correlations between the first two components of the EFAWS and the ratings that the FA assessors gave to 14 welfare issues are shown in Table 3. There were significant positive correlations between the FA assessors' values for component 1 (pigs have mental welfare) and how much of a problem they rated the following welfare issues for pigs: boredom ($r_s = 0.530$, $P = 0.006$), fighting ($r_s = 0.430$, $P = 0.032$), bursae ($r_s = 0.470$, $P = 0.016$) and lack of a manipulable material but presence of an enrichment object, such as a chain ($r_s = 0.408$, $P = 0.042$). FA assessors' values for component 2 (people-centric, pigs as profit) was only found to correlate significantly with their ratings of how important fear of people, including the stockman is ($r_s = 0.420$, $P = 0.044$).

Discussion

Although this is the largest number of Farm Assurance assessors reported to be formally trained in welfare outcome assessments, the relatively small sample size means that careful statistical interpretation is required. It is more usual for Principal Component Analysis to be conducted on a larger number of respondents, with a rule of thumb being ten times the number of respondents as questions or at least 100 respondents. However, a small sample size does not invalidate the procedure, it just makes it less likely that the data will generalise into reliable components. If the data does generalise into components which have high loading items (~ 0.8) then it is concluded the number of respondents is adequate for the procedure (Tabachnick & Fidell 2007). The first two components produced following PCA of the ECAWS ('pigs have mental welfare' and 'people-centric, pigs as profit') both had high loading items (-0.795 and 0.808 , respectively) and were similar to the two 'higher-order' components, termed 'welfare' and 'business orientation', found by Austin *et al* (2005) when they tested the EFAWS on sheep and pig farmers.

The results of the different tasks indicate that there is variation between assessors, between welfare outcome measures and between test settings (ie individual pigs, photographs or prevalence in pens of pigs) following training, giving rise to differences in levels of agreement with the gold standard. A good level of agreement with the gold standard, suitable for use in Farm Assurance schemes, was not reached by all assessors for all measures after this short period of training. It would appear that further training sessions would be required to achieve good levels of agreement for some assessors, although the extent of training required, and whether all assessors are able to achieve good levels of agreement, was not tested in this study. The relatively few (7 out of a possible 98) correlations between the two major principal components and measures of agreement between assessors suggest that assessing these welfare outcomes to a defined standard protocol is reasonably resistant to bias as a result of difference in observer attitude to farm animal welfare. It could be expected that between four and five correlations may occur

by chance alone. However, in those tests where the number of assessors was reduced, the likelihood of achieving a significant correlation will have been reduced.

The correlations that were significant mainly related to component 1 ('pigs have mental welfare') and 'pigs requiring hospitalisation'. Those FA assessors that had higher scores for component 1 also recorded more pigs as requiring hospitalisation. There could be two reasons for this. Firstly, these assessors may be more likely to ascribe a negative mental welfare state to a sick or injured pig and therefore believe that they would benefit from transfer to the more nurturing environment of a hospital pen. Secondly, this welfare outcome measure, of all the measures FA assessors were trained in, had the largest degree of judgement required when making the assessment and therefore may be more open to the influence of attitude compared to other, more tightly defined, measures. That this difference in prevalence scores did not result in a significant worsening of agreement with a gold standard is probably because the gold standard trainer had a prevalence score somewhere in the middle of the overall range of prevalences scored by all FA assessors. Therefore, there will be less difference (better agreement) between the gold standard trainer and any one FA assessor than between the FA assessors at the extreme ends of the range of prevalence scores. This result suggests that care should be taken to ensure that any new measures proposed for use for welfare assurance should be tightly defined and aim to rely little on judgment in order to improve the likelihood of agreement between observers, whatever their underlying attitude to farm animal welfare.

There were also significant correlations between component 1 and the recorded prevalence of tail lesions and lameness in one of the three training sessions, although not in prevalence residuals or levels of agreement. The significance of this result is unknown as it was not replicated in the other training sessions and a larger sample size may be useful to determine whether this is a real effect.

The FA assessors that had higher scores for component 2 (people-centric, pigs as profit) recorded more pigs with tail lesions, but fewer pigs with severe tail lesions, in one training. This suggests that a different, less sensitive, threshold for severity was adopted by these subjects. Milder tail lesions may give more information about all the pigs in the pen whereas severe tail lesions in our protocol are more significant to the individual pig as they will be painful.

The only effect that attitude was found to have on statistical levels of agreement was a negative correlation between those FA assessors that had higher scores for component 2 and their agreement with a gold standard for lameness. However, there was an absence of any significant difference of recorded prevalence of lameness and as lameness is one of the more difficult outcome measures to standardise this result must be treated with caution until further verification is obtained.

The ratings that FA assessors gave to a range of welfare problems highlighted that they perceived being tail bitten, not being removed to a hospital pen if sick/injured and having body lesions or lameness as being the most problematic for finishing pigs. The lowest ratings were for dirtiness, having a manipulable object but no manipulable material,

poor growth rate and boredom. When a panel of experts were asked to identify the most appropriate measures for welfare assessment on pig farms they cited observing lameness and examining limb lesions and mortality and medicine records as being most important (Whay *et al* 2003). In our study, the FA assessors were not given a free choice of response but were only able to rate those measures presented to them. The significant positive correlations between the FA assessors scores for component 1 ('pigs have mental welfare') and welfare ratings for boredom, fighting, bursae and having a manipulable object but no manipulable material may result from identification of those areas where there is a clear mental component to the problem without necessarily a physical one. The FA assessors with higher component 2 scores ('people-centric, pigs as profit') appear to be more concerned about the relationship between pigs and people on the farm, including stockmen.

Animal welfare implications and conclusion

Accurate assessments of animal welfare are often the first step to improving animal welfare. The Farm Assurance system for pigs in the UK could potentially be a useful mechanism for delivering welfare assurance of, and eventually welfare improvement to, a large number of pigs. In order for this to occur, the formal inclusion of welfare outcome measures into FA schemes must result in the production of reliable data. One aspect of this is ensuring consistency between assessors, so that there is no bias depending on which assessor visits a farm. Farm Assurance assessors in this study had a range in attitudes to farm animal welfare. They also had a range in both the prevalence of welfare outcome measures that they recorded and agreement with a gold standard following training. Despite this, our results indicate that training in welfare outcomes, defined by a standard protocol, is relatively unconfounded by a difference in observer attitude farm animal welfare. To obtain better levels of agreement between assessors, and therefore more reliable data, it is recommended that FA schemes concentrate their resources on providing good quality training in a well-defined protocol and reliability testing and that they do not need to attempt to account for the attitudes of the FA assessors to farm animal welfare.

Acknowledgements

This study was funded by the British Pig Executive and we thank them, the FA assessors and the pig farmers who allowed the training to take place on their farms, for their co-operation.

References

- ABP** 2007 *Assured British Pigs Certification Standards for Pigs*. ABP: Cobham, UK
- Austin EJ, Deary IJ, Edwards-Jones G and Arey D** 2005 Attitudes to farm animal welfare. Factor structure and personality correlates in farmers and agriculture students. *Journal of Individual Differences* 26: 107-120
- FAWC** 2005 *Report on the Welfare Implications of Farm Assurance Schemes*. Farm Animal Welfare Council: London, UK
- GQA** 2007 *Genesis Quality Assurance Certification Standards for Pigs*. GQA: Burton-on-Trent, UK
- Keeling L** 2007 Turning welfare principles into practice: approach taken by welfare quality. In: Veissier I, Forkman B and Jones B (eds) *Assuring Animal Welfare: From Societal Concerns to Implementation* pp 25-28. Berlin, Germany
- Kristensen E, Dueholm L, Vink D, Andersen JE, Jakobsen EB, Illum-Nielsen S, Petersen FA and Enevoldsen C** 2006 Within- and across-person uniformity of body condition scoring in Danish holstein cattle. *Journal of Dairy Science* 89: 3721-3728
- Main DCJ and Green LE** 2000 Descriptive analysis of the operation of the Farm Assured British Pigs scheme. *Veterinary Record* 147: 162-163
- Main DCJ, Clegg J, Spatz A and Green LE** 2000 Repeatability of a lameness scoring system for finishing pigs. *Veterinary Record* 147: 574-576
- Main DCJ, Whay HR, Leeb C and Webster AJF** 2007 Formal animal-based welfare assessment in UK certification schemes. *Animal Welfare* 16: 233-236
- March S, Brinkmann J and Winkler C** 2007 Effect of training on the inter-observer reliability of lameness scoring in dairy cattle. *Animal Welfare* 16: 131-133
- Mullan S, Browne WJ, Edwards S, Butterworth A, Whay HR and Main DCJ** 2009 The effect of sampling strategy on the estimated prevalence of welfare outcome measures on finishing pig farms. *Applied Animal Behaviour Science* 119: 39-48
- Mullan S, Edwards S, Whay HR, Butterworth A and Main D** 2011 Interdependence of welfare outcome measures and potential confounding factors on finishing pig farms. *Applied Animal Behaviour Science* 121: 25-31
- Petersen HH, Enoe C and Nielsen EO** 2004 Observer agreement on pen level prevalence of clinical signs in finishing pigs. *Preventive Veterinary Medicine* 64: 147-156
- Shaw P** 2003. *Multivariate Statistics for the Environmental Sciences*. Hodder Arnold: Oxford University Press: New York, USA
- Tabachnick BG and Fidell LS** 2007 *Using Multivariate Statistics, Fifth Edition*. Pearson: Boston, USA
- The Welfare of Farmed Animals (England) Regulations** 2007 *The Welfare of Farmed Animals (England) Regulations*. Office of Public Sector Information: London, UK
- Whay HR, Main DCJ, Green LE and Webster AJF** 2003 Animal-based measures for the assessment of welfare state of dairy cattle, pigs and laying hens: consensus of expert opinion. *Animal Welfare* 12: 205-217