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

Animal Welfare 2012, 21: 389-396
ISSN 0962-7286
doi: 10.7120/09627286.21.3.389

Welfare outcomes assessment in laying hen farm assurance schemes

DCJ Main*†, S Mullan†, C Atkinson‡, A Bond‡, M Cooper§, A Fraser§ and WJ Browne†

† University of Bristol Veterinary School, Langford BS40 5DU, UK
‡ Soil Association, South Plaza, Marlborough Street, Bristol BS1 3NX, UK
§ Royal Society for the Prevention of Cruelty to Animals, Wilberforce Way, Southwater, Horsham, West Sussex RH13 9RS, UK
* Contact for correspondence and requests for reprints: d.c.j.main@bristol.ac.uk

Abstract

Most farm assurance schemes in the UK at least, in part, aim to provide assurances to consumers and retailers of compliance with welfare standards. Inclusion of welfare outcome assessments into the relevant inspection procedures provides a mechanism to improve animal welfare within assurance schemes. In this study, taking laying hens as an example, we describe a process for dealing with the practical difficulties in achieving this in two UK schemes; Freedom Food and Soil Association. The key challenges arise from selecting the most appropriate measures, defining sampling strategies that are feasible and robust, ensuring assessors can deliver a consistent evaluation and establishing a mechanism to achieve positive change. After a consultation exercise and pilot study, five measures (feather cover, cleanliness, aggressive behaviour, management of sick or injured birds, and beak trimming) were included within the inspection procedures of the schemes. The chosen sampling strategy of assessing 50 birds without handling provided reasonable certainty at a scheme level but less certainty at an individual farm level. Despite the inherent limitations within a time and cost sensitive certification assessment, the approach adopted does provide a foundation for welfare improvement by being able to highlight areas of concern requiring attention, enabling schemes to promote the use of outcome scoring as a management tool, promoting the dissemination of relevant technical information in a timely manner and increasing the scrutiny of standards important for the welfare of the birds.

Keywords: *animal welfare, farm assurance, feather loss, laying hen, sampling, welfare outcomes*

Introduction

In general, farm assurance schemes aim to provide assurances to the consumer and the foodchain on compliance with food safety, animal welfare and environmental standards. These private standards are usually developed by reference to existing legislation, codes of practice, scientific knowledge and practical experience. Farm assurance standards or marketplace requirements have been primarily focused on resource (engineering) rather than outcome (performance) standards (Mench 2003; Webster 2009). However, the increased inclusion of welfare outcomes within farm assurance schemes has previously been advocated by the Farm Animal Welfare Council (FAWC 2005). The Farm Animal Welfare Forum (2010) has also advocated that welfare outcome 'safeguards' be used to provide evidence about the 'welfare credentials' of production system labelling.

A collaborative project (AssureWel) between the University of Bristol, the Royal Society for the Prevention of Cruelty to Animals (RSPCA) and Soil Association aimed to introduce a formalised and structured approach to including welfare outcomes into the inspection procedures of the Freedom Food and Soil Association Certification schemes, initially focusing on laying hens. The Freedom Food scheme, which is owned by the RSPCA, has over 1,000 members with laying hen farms and accounts for approximately 99% of the non-cage and approximately 50% of the total egg production in the UK (RSPCA, personal communication 2011). The Soil Association scheme which also incorporates the relevant organic regulation (European Community 2007, 2008) has over 250 members with laying hen farms. Although both schemes specify many detailed resource requirements, the use of welfare outcomes has not yet been fully developed in a structured and formalised way to its full potential. For example, both schemes' standards make reference to feather loss, a welfare outcome measure. An assurance assessor verifying compliance with this standard would need to evaluate, even if only informally, the level of feather loss to determine whether a behavioural problem had occurred. However, it is recognised that assessment could be improved with a more scientifically robust methodology. Previous research investigations (Green *et al* 2000; Bestman & Wagenaar 2003; Whay *et al* 2007; Lambton *et al* 2010) have shown significant variability in the levels of injurious pecking in UK non-caged hens. For schemes to be able to

promote a further improvement in bird welfare, the inspection procedures need to be able to accurately monitor and encourage the pro-active management of outcome measures, such as injurious pecking.

The general and now widespread interest in welfare outcomes has culminated in the formulation of standardised welfare assessment protocols for poultry, cattle and pigs (Welfare Quality® 2009a,b,c). These protocols provide a mechanism for evaluating farms against 12 welfare criteria and four overarching principles to produce a single overall welfare score for a farm. These protocols could be used for research, certification, management and legislative purposes (Main *et al* 2003). However, for use with farm assurance certification schemes, the time needed for the full Welfare Quality® evaluation, normally several hours on most farms, currently limits its potential application in UK farm assurance schemes. Extending farm assurance assessments by more than approximately 30 min could affect the number of visits that could be undertaken in one day. However, there is some scope for the development and inclusion of assessment methodologies to support the assessment of existing outcome-based standards for both the RSPCA Freedom Food and Soil Association assurance schemes.

This study outlines the process for selecting the most appropriate welfare outcome measures, defining the optimum sampling strategy, ensuring assessors were able to consistently assess outcomes and providing suitable feedback on levels of outcomes to farmers. The specific aims were to improve the assessment of standards that currently require an informal assessment of outcomes, to promote interest in the use of outcome measures as management tools at a farm level and to provide information at a scheme level on welfare outcomes, ie facilitate more evidence-based decisions with future standards.

Materials and methods

Selection of measures

A list of possible measures was drawn up by reviewing existing standards of both schemes, previous welfare assessment protocols (Whay *et al* 2007; Welfare Quality® 2009b) and relevant research investigations. Specific protocols for each measure were then formulated and discussed in detail with the technical experts working within the Freedom Food and Soil Association schemes, welfare scientists within the Bristol Animal Welfare and Behaviour group and laying hen producers. The measures and assessment methodologies were then trialled as part of a pilot study on a small number of laying hen farms considered representative of the range of farms existing on each scheme. After the pilot study further amendments were made to the detailed assessment guidelines.

The following principles were used to select the potential welfare measures. The measures should both improve the assessment of existing standards (RSPCA 2011; Soil Association 2011) and relate to an aspect of husbandry or care that had a significant welfare impact. Since the measures were not intended to be aggregated into a single score, they did not need to include measures for each of the five freedoms (FAWC 2005) or each of the 12 Welfare Quality® criteria (Botreau *et al* 2007). Where possible, the assessment methods were equivalent to the relevant Welfare Quality® parameter to enable future comparisons with other datasets.

Once finalised, the detailed assessment protocols were pilot tested on ten laying-hen farms. The pilot consisted of the assessment of 100 birds for feather loss over the back and neck of the bird, dirtiness, beak length and shape, comb-pecking marks, skin lesions on the rest of the body and management of sick/injured birds. In addition, the results of these assessments were discussed with the farmer and the responses of both the farmers and assessor to the process were later distilled by each assessor in an informal report of their visits.

Sampling methods

Ensuring an appropriate level of accuracy, whilst ensuring the total time was feasible within the existing assessment process, was a key concern for the schemes. Therefore, in order to define the most appropriate sampling strategy for farm assurance, information was needed on the accuracy of prevalence estimates from assessment of a variety of sample sizes of birds on-farm. The welfare outcome data had the potential to be used at the 'farm level' for benchmarking or certification decisions and at 'scheme level' for overall scheme monitoring. The accuracy of prevalence estimates for both farm and scheme level was, therefore, assessed.

The 95% confidence interval widths were calculated for farm level prevalence estimates of outcome present versus absent measures (that took values between 0 and 100%) for sample sizes of 50, 100, 200 and 1,000 birds from an assumed infinite population, and for 100 birds from a population of 2,000 birds. The following statistical formula was used to calculate the 95% confidence interval width (four times the standard error) of the prevalence assessed in a sample of randomly selected birds from an infinite population:

Infinite population 95% confidence interval (CI) = $4 \times sqrt(p(1\text{-}p)/n)$

where p was the prevalence recorded from the sample of n birds.

In practice, at any one time, a flock has a fixed size and flock size will vary both within and between schemes. This calculation for an infinite population reduces with increasing sample size but clearly with a finite population we will know exactly the proportion if we sample all birds and so an adjustment is required. A finite population correction (FPC) factor was therefore also calculated to be able to account for the difference between these actual population sizes and an infinite population through the following formula:

Finite population correction (FPC) = $sqrt([N\text{-}n]/[N\text{-}1])$

where N was the total flock size from which the sample of n birds are assessed.

The finite population confidence interval width (95% CI) could then be calculated as follows:

Finite population 95%CI = FPC × infinite population 95% CI

https://doi.org/10.7120/09627286.21.3.389 Published online by Cambridge University Press

Calculations were also performed to estimate the accuracy of an overall farm assurance scheme level prevalence. For these calculations, the sample size was the total number of birds assessed (the number of birds assessed per farm, multiplied by the number of farms in that scheme, ie 1,067 Freedom Food farms and 250 Soil Association farms). A finite population correction factor was not used on this occasion because it would make minimal difference given the large number of hens involved in a scheme.

Assessor training and feedback to producers

In order to ensure assessors were consistent in their approach, assessors from Soil Association Certification and the Freedom Food scheme, field staff from the RSPCA Farm Animals Department, and eight employees from a laying-hen producer company were trained in the five welfare outcome measures selected based on results of the pilot study. The training sessions were organised on three free-range laying hen units. Free-range systems were used as these provided the most complex system for assessment, and neither scheme permit the use of caged systems. The training started with photographic familiarisation with the non-behavioural measures and progressed to discussion of live bird assessments in small groups with an experienced observer. Finally, each assessor undertook a repeatability test by recording their scores for feather loss in two body regions and dirtiness for 50 live birds. This live bird repeatability testing occurred in ten groups of between five and eight observers to ensure each observer could assess the same birds. The groups of assessors also discussed the other parameters (aggressive behaviours, sick and injured and beak trimming) whilst observing the birds. As no-one was considered sufficiently experienced to be a gold standard, the prevalence data were analysed for closeness of prevalence to other group members and agreement with the group mode, ie the most common response for each individual bird.

In addition to the training days, an online tool was developed for feather loss and cleanliness which assessors completed following their initial training. It consisted of three stages: i) a series of photographs depicting birds of each score with a written description; ii) a short training test where immediate feedback against a pre-defined gold standard was available after assessors scored each of 15 photographs; and iii) a test of 50 photographs, without feedback in between, but with feedback on their percentage agreement with the gold standard (as defined within the Welfare Quality® training material) at the end.

The discussion with the farmer about their welfare outcome results was believed key to stimulating the farmer to make changes resulting in welfare improvements. As the schemes are also accredited against the ISO Guide 65 /EN45011 standard it was important not to compromise the impartiality of the assessment. The nature of the type of discussion that could take place was, therefore, clarified with the United Kingdom Accreditation Service (UKAS) to ensure that impartiality was not breached by assessors. Guidance on maintaining impartiality whilst promoting further interest in welfare assessment and improvement was included in the training day for the farm assurance assessors.

Results

Measures

The final outcome measures selected for inclusion in the Freedom Food and Soil Association schemes were feather cover, aggressive behaviours, beak trimming, bird dirtiness and the management of sick or injured birds (Table 1). The intention was that the assessment protocols were equivalent to measures within the Welfare Quality® protocol (2009b). However, there were significant limitations in a farm assurance context. For example, during the pilot exercise it was clear that assessors were not able to catch individual birds to evaluate issues such as feather loss and keel-bone fractures as described in the Welfare Quality® (2009b) protocol. Picking up birds was considered to be too time consuming, not possible or likely to induce a considerable flight response in some flocks. This meant that feather loss was assessed by observing the birds. Also, not being able to hold birds meant that assessing keel-bone fractures was not possible at this stage of the project. In addition, the project team considered that the uncertainty over the meaningfulness, reliability, feasibility or time-efficiency meant that other measures (such as comb pecking marks and skin lesions and abnormal beak shape) were removed as stand-alone measures from formal monitoring. However, when severe, these issues were included as potential causes of sick or injured birds. After the pilot phase, a whole flock assessment of aggressive behaviour was added in place of comb pecking marks. The aim here was to ensure that aggressive behaviour, which is distinct from injurious feather pecking behaviour (see, for example, Nicol *et al* 1999), was included as described in the Welfare Quality® protocol (2009b).

Sampling methods

The 95% confidence interval widths for different estimates of farm level prevalence and using different sample sizes are shown in Figure 1. For a sample of 100 birds on a 2,000 bird flock (ie the maximum number for a Soil Association certified unit) the finite population correction factor was 0.975. For a sample of 100 birds on a 16,000 bird flock (ie maximum number for a Freedom Food certified unit) the finite population correction factor was 0.997. These correction factors were considered close enough to the infinite population estimate as to not make a difference in practice (Figure 1).

An estimate of the maximum 95% confidence interval width for a sample of 12,500 birds taken randomly from all Soil Association farms (equivalent of 50 birds from each of the 250 farms) and 53,350 birds from Freedom Food farms (equivalent of 50 birds from each of the 1,067 farms) was 1.8 and 0.9%, respectively. This statistical approach assumes random sampling of all birds across the scheme regardless of which farm they are on. In reality, a stratified approach would be used where birds are sampled from every farm in the scheme and it is anticipated that this would provide a greater level of accuracy in the scheme level prevalence, particularly if the average farm prevalence is required. However, to

Table 1 Measures selected for inclusion within the Freedom Food and Soil Association scheme assessment procedures.

Measure	Protocol
Feather cover* (quantitative measure)	Assess and score 5 randomly selected birds in each of 10 different areas of the house and/or range. Visually assess the head and neck area and back and rump area of the bird. Score separately for head and neck area, and back and rump area. Score 0: No/minimal feather loss. No bare skin visible, no or slight wear, only single feathers missing. Score 1 : Slight feather loss. Moderate wear, damaged feathers or 2 or more adjacent feathers missing up to bare skin visible < 5 cm maximum dimension. Score 2: Moderate/severe feather loss. Bare skin visible ≥ 5 cm maximum dimension.
Bird cleanliness* (quantitative measure)	Assess and score 5 randomly selected birds in each of 10 different areas of the house and/or range. Visual assessment of one side of the bird, except the feet and legs. Score 0: Clean. The bird is clean. Score 1: Moderate dirtiness. There is soiling on at least one part of the bird but no area ≥ 5 cm maximum dimension. Score 2: Substantial dirtiness. There is soiling on one or more parts of the bird ≥ 5 cm maximum dimension.
Beak trimming (semi-quantitative measure)	Refer to chick placement records and/or ask the unit manager to determine whether/when the birds were either: 1) not beak trimmed, 2) beak trimmed before 10 days of age or 3) beak trimmed as emergency procedure under veterinary advice. Visually assess birds' beaks during the assessment time spent in the house and range. Record the number of any birds seen which have more than 1/3 of their beak removed.
Aggressive behaviours* (semi-quantitative measure)	Observe and listen to the behaviour of birds during the assessment time spent in the house and range. Aggressive behaviour is defined as fighting, severe pecking at other birds or chasing other birds (when observed more than twice). Aggressive behaviours are often signalled by a loud squawk or vocalisation. Record number of incidents of aggressive behaviour observed or heard during your total time spent with the flock (in the shed or on the range).
Management of sick or injured birds (semi-quantitative measure)	Visually assess birds during the assessment time in the house and range to identify sick or injured birds that would benefit from hospitalisation (ie removal from the main flock) or should be euthanased. This would include obviously sick birds (with fluffed up feathers and an inactive, unresponsive appearance) and birds with body wounds that have fresh blood that might attract cannibalistic attention from other birds. Include birds in hospital pen that should be euthanased.

* Adapted from Welfare Quality® (2009).

improve the accuracy of an estimate of overall scheme level prevalence (proportion of affected birds in the whole scheme) a sampling strategy weighted for farm size would be preferable to avoid the potential for bias if the measure was correlated with farm size.

Discussions between the AssureWel project partners on the trade-off between the time constraints of conducting the assessment and the usefulness of the information gained from any particular sample size resulted in agreement of a sample size of 50 birds to be assessed per farm. This would give useful information about the whole farm assurance scheme but less accurate information about individual farms, limiting its value for farm level comparisons.

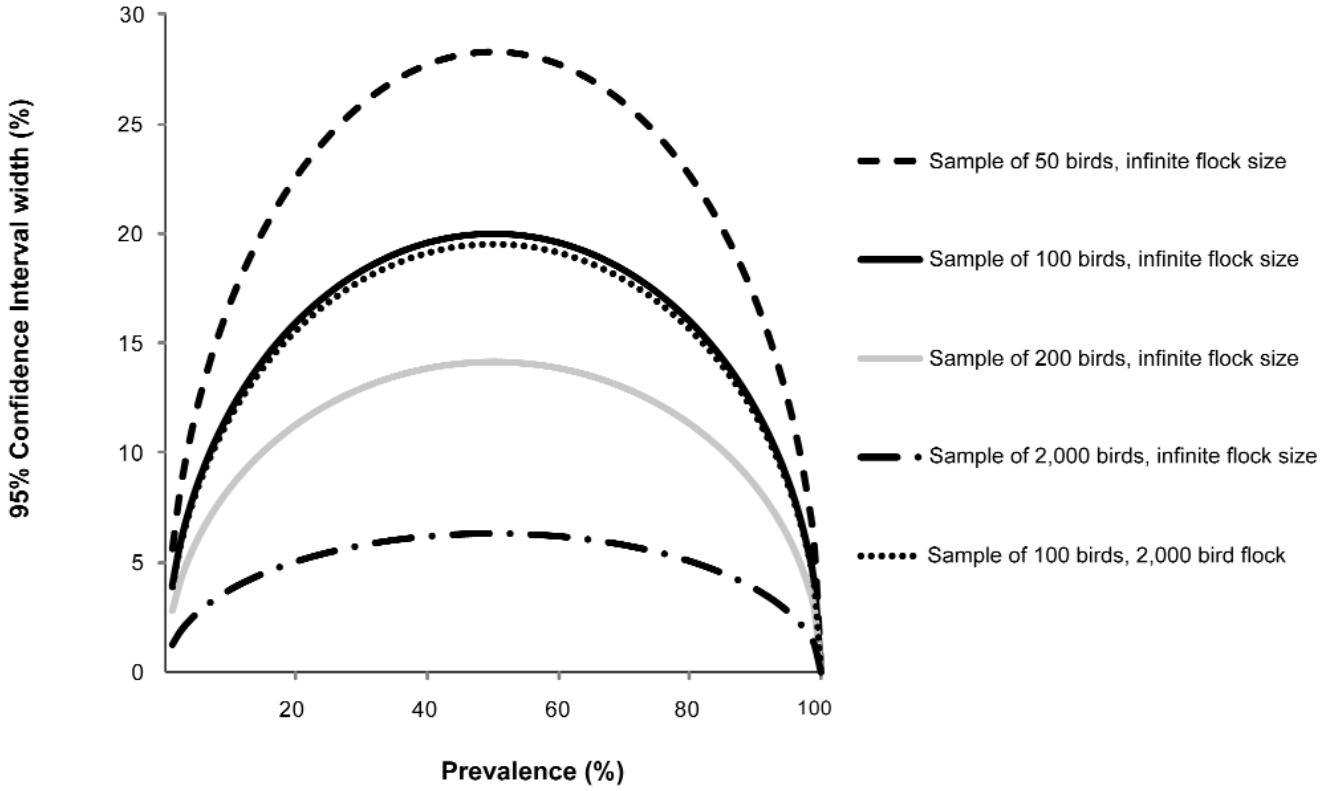
Assessor training and feedback to producers

As an example of the repeatability testing procedure, the results for feather cover are described. The prevalences of any back or rump feather loss observed by a total of 64 assessors and field staff working in ten groups with each group observing 50 different birds are shown in Figure 2. The overall farm prevalence is the unit of interest, rather than identification of individual animals with specific scores (Mullan *et al* 2011). For feather loss, the average range in prevalence reported by the assessors observing each set of 50 birds was 21.8% ranging from 8 to 44%. Out of the 64 observers, 43 had more than 80% agreement with the group mode, ie the most common response for each individual bird. After 2 months access to the online tool, eleven out of 56 farm assurance assessors and RSPCA field staff had completed all three stages and eight out of the eleven assessors achieved 80% agreement with the gold standard scores.

With regards to the feedback of results to the producer, the keypoints of guidance provided to assessors are summarised in Table 2. With respect to the conversation with the farmer about their farm's outcome results, the view of UKAS was that farm assurance assessors could discuss all aspects of a farm's welfare problems and solutions as long as farm-specific advice was not offered. Information sources that cite generic solutions to a welfare problem, such as leaflets or websites, could, therefore, be offered and further support, such as seeking veterinary advice, could be encouraged. As the RSPCA and Soil Association standards contain sufficient standards that can be linked to feather loss, discussing this with the producer was identified by UKAS as being acceptable and considered useful by assessors in stimulating a meaningful conversation with the farmer. If the assessor identified a particular problem after observing the animals, the assessor was encouraged to examine in detail compliance

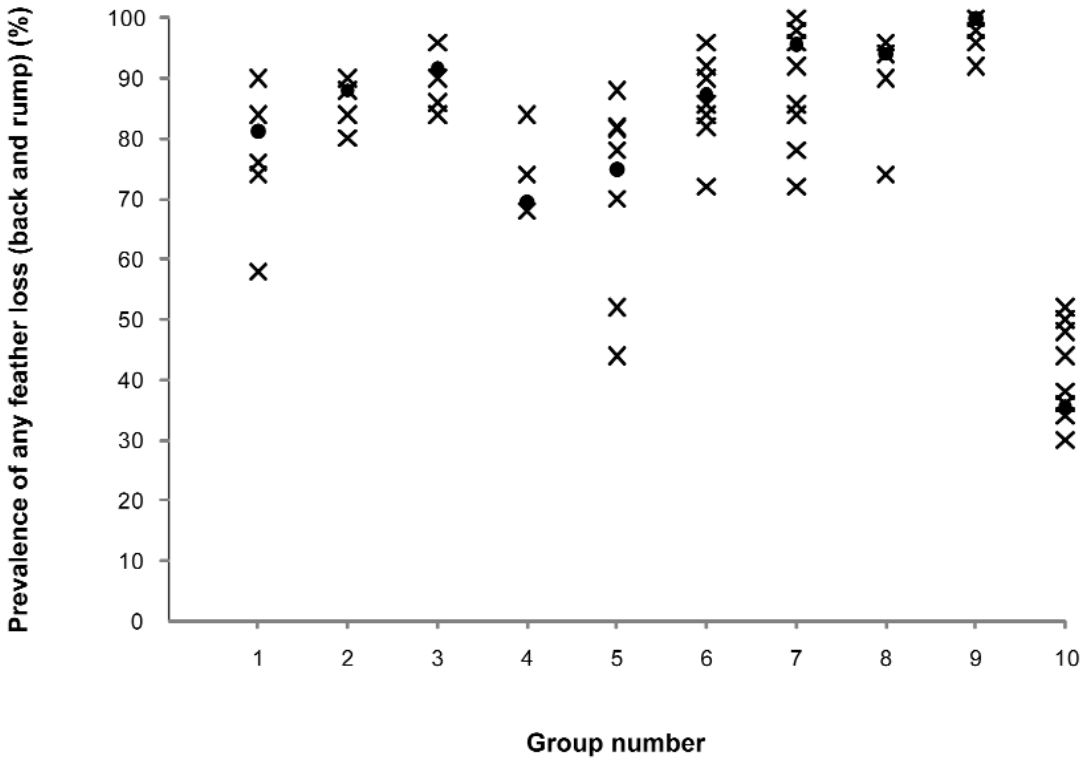
https://doi.org/10.7120/09627286.21.3.389 Published online by Cambridge University Press

Figure 1



The 95% confidence intervals of farm level prevalence estimates derived from a range of sample sizes.

Figure 2



The prevalence of feather loss recorded by 64 assessors (**X**) working in one of 10 groups and their group mode (•) after an on-farm training exercise from 3 laying hen farms. Groups numbered 1–4, 5–9 were undertaken on the same farm. For each group a different set of 50 birds were observed and each assessor was asked to record the presence of 'any feather loss on either the back or the rump'. The group mode was the prevalence within each set of 50 birds based on the most common response for each individual bird.

Table 2 Keypoints of guidance provided to farm assurance assessors on a) the scope of conversations with farmers, b) the tactics for promoting behaviour change and, using feather cover as example, c) non-farm specific technical advice that can be provided on potential solutions and d) the relevant scheme standards that would need further investigation if a potential problem was observed.

Type of guidance	Keypoints/summary of content
a) Nature of conversation with farmers in order to remain compliant with EN 45011	Assessors must not: • Give specific prescriptive advice • Provide instruction on possible solutions Assessors can discuss the following: • Justification of the standard • Description of a problem • Encourage interest and awareness of the problem • Refer farmer to other advisors • Refer farmer to information sources • Provide technical guides approved by the scheme • Advocate the value of advice (from others) • Explain that other farmers have solved this problem • Explain benefits of solving the problem
b) Possible tactics for encouraging behaviour change as determined during farm assurance assessor workshop	Major themes for discussion: • Positive encouragement and understanding from the assessor • A knowledgeable and competent assessor • Avoiding confrontation with farmer • Show interest in farm situation
c) Generic advice on solutions for feather loss	The following areas are covered in leaflets provided to farmer (available from following website www.assurewel.org): • Causes of the welfare issue (eg low health status pullets) • Avoiding the welfare issue (eg ensure pullet placement is optimal) • Managing the welfare issues (eg improve enrichment) • Sources of further information
d) Examples of scheme standards and government guidance that require observation of feather loss (other standards may also be relevant)	• RSPCA standard H1.1: The written Veterinary Health and Welfare Plan must "set targets for health aspects" (including significant feather loss) and "record whether those targets have been met each year" • Soil Association: There is an allowance for a larger flock size if "you can show us that you can maintain a high level of bird health and welfare". • Defra code of recommendations (Defra 2002): If behavioural problems occur, which manifest themselves in injurious feather pecking, they should be tackled immediately by appropriate changes in the system of management

with the relevant standards. In most cases, the assessor would be expected to verify that the appropriate flock level corrective action and individual animal treatments were provided to the animals. At this initial stage, more formalised intervention guidelines were not introduced particularly as sufficient scheme level data had not been collected.

Discussion

This study describes the process undertaken to develop a practical and achievable mechanism for embedding laying-hen welfare outcome assessment into two schemes: RSPCA Freedom Food and Soil Association. This study utilised a mixture of quantitative methods, such as sampling and repeatability tests, and more qualitative approaches using consultation exercises and a review of standards, to determine the optimum methods of including welfare outcomes within the assurance schemes. The limitations arising from the practical application to farm assurance schemes operating in a commercial environment are quite clear. In particular, the limited time available without incurring additional cost to farmers restricted the numbers of animals that could be reasonably assessed to 50 birds for two parameters each and at flock level for three parameters. Despite the limitations in the approach the inclusion of some outcomes does represent significant opportunities to improve welfare within both schemes. Once the assessments have been undertaken on more than 200 units, after three months, the protocols will be further reviewed and adapted as necessary.

Limitations

The pilot visits were invaluable in identifying the practicality and usefulness of some of the measures, as well as gaining an understanding of the acceptability of such assessments to the farmer. The strong desire for a robust assurance process meant that some measures were removed early from formal assessment. When considering application to other species, it may be valuable to include a larger number of formal outcome measures at the initial stages with the expectation that some measures will not

https://doi.org/10.7120/09627286.21.3.389 Published online by Cambridge University Press

prove suitable. In addition, more than one pilot phase may help deal with some of the feasibility concerns. An example of one measure that was excluded at an early stage was keel-bone damage (Wilkins *et al* 2004) which the Farm Animal Welfare Council (2010) has identified as an important welfare concern in laying-hen flocks. The technique was not included because of difficulties in picking up birds. Subsequent reviews of the welfare outcome protocols used by each scheme will keep this issue under consideration. It is hoped that alternative methods to encourage farmer awareness and interest in fractures will be included whenever possible.

Defining optimum sample size will always remain a difficult decision. The relatively narrow confidence interval and hence high level of confidence of scheme level data (ie less than 2%) arising from a 50-bird evaluation per farm does mean that data should be available in due course for self assessment and benchmarking where a larger number of birds could be assessed. The scheme-level data will also be useful for scheme-level policies, such as further development of the standards. However, the 50-bird sample size is a significant limitation for farm level decisions. Based on the theoretical modelling of sample size, this meant that the farm-level confidence interval width could be as high as 25–30%. Farm-level data should therefore be used with caution with a clear explanation of the potential variability in the result. This will be a particularly important consideration in deciding how the schemes can introduce formalised benchmarking. The uncertainty at the farm level does mean that setting an intervention guideline above which a non-compliance should be issued is unlikely to be possible with this sampling approach. However, sampling more animals on a risk basis may enable the use of more satisfactory and robust intervention guidelines in the future.

Finally, despite on-farm standardisation, there were some variations in the assessment between different assessors. Some variation is inevitable with animal-based scoring systems and has been reported in other scoring systems (Mullan *et al* 2011). A binary response was used for repeatability assessment as a clear present or absent response is useful for policy discussions. Other authors, such as Brenninkmeyer *et al* (2007) and Channon *et al* (2009), have used binary levels of repeatability when data were collected on a more complex scale.

Assessors are already required to make 'judgements' concerning compliance with scheme standards based on their previous experience and interpretation of guidance notes. These professional judgements have not been subject to scrutiny in terms of repeatability testing. The inclusion of formal outcomes now provides an opportunity to measure and, if necessary, improve such judgements. The online tool was also available after the initial training which should help assessors maintain consistent scoring. It is hoped that this will be a more cost-effective approach to enable easily accessible and regular refresher training than frequent on-farm training days. However, it is anticipated that some regular on-farm standardisation may also be required.

Welfare improvement

Despite the limitations discussed, the inclusion of welfare outcomes within assurance schemes does provide opportunities to improve welfare. For example, Main and Mullan (2012) have previously argued that farm assurance schemes could encourage active participation and interest in both outcome scoring systems and possible husbandry solutions. The approach described here for laying hens can promote welfare improvement in several specific ways. Firstly, the discussion that takes place between the farm assurance assessor and farmer is an opportunity to encourage an active interest in using a formalised scoring system and the welfare issues assessed. It is commonly reported by both assessors and farmers that they would like to spend more time observing and talking about the animals rather than evaluating the adequacy of farm documents and records. Secondly, whilst assessors are restricted from providing farm-specific advice about a single solution, they can provide generic detailed technical information on particular issues, they can make the farmer aware of and promote the value of other sources of advice and more generally they can, based on personal experiences, explain that others have addressed these issues. The assessor does, however, need to make it clear from the outset that they cannot provide farm-specific advice that would compromise the impartiality of future assessments in accordance with requirements governing the accreditation of the schemes (EN45011/ISO Guide 65). Thirdly, the scheme-level results could be used as a benchmarking management tool if sufficient data are collected, which is likely to motivate positive changes to improve welfare. Finally, the increase in animal observation time is likely to increase the scrutiny of standards that are important for the birds, ie it is hoped that standards which were previously difficult to assess, but important for the animal, receive more attention from the assessor.

In summary, the approach presented should help schemes improve welfare by highlighting areas of concern requiring attention, promoting the use of outcome scoring as a management tool and encouraging the dissemination of relevant technical information.

Acknowledgements

The AssureWel project is a collaboration between University of Bristol, RSPCA and Soil Association that was made possible with support from Tubney Charitable Trust. The project team actively involved in developing this process were Iain Rogerson, Kate Still, Madeleine Neale, Alice Clark, Allan Pearson, Julia Wrathall, Bob Waller, Dorian Cross and Leigh Grant. We are particularly grateful to the pilot farms and Stonegate for active participation in the pilot phase.

References

Bestman MWP and Wagenaar JP 2003 Farm level factors associated with feather pecking in organic laying hens. *Livestock Production Science 80*: 133-140. http://dx.doi.org/10.1016/S0301-6226(02)00314-7

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

Brenninkmeyer C, Dippel S, March S, Brinkmann J, Winckler C and Knierim U 2007 Reliability of a subjective lameness scoring system for dairy cows. *Animal Welfare 16*: 127-129

Channon A, Walker A, Pfau T, Sheldon I and Wilson A 2009 Variability of Manson and Leaver locomotion scores assigned to dairy cows by different observers. *The Veterinary Record 164*: 388-392

European Community 2007 Council Regulation (EC) No 834/2007 of 28 June 2007 on organic production and labelling of organic products and repealing Regulation (EEC) No 2092/91. *Official Journal L189*: 1-23. http://dx.doi.org/10.1136/vr.164.13.388

European Community 2008 Commission Regulation (EC) No 889/2008 of 5 September 2008 laying down detailed rules for the implementation of Council Regulation (EC) No 834/2007 on organic production and labelling of organic products with regard to organic production, labelling and control. *Official Journal L250*: 1-84

FAWC 2005 *Report on welfare implications of farm assurance schemes.* Farm Animal Welfare Council: London, UK

FAWC 2010 *Opinion on osteoporosis and bone fractures in laying hens.* Farm Animal Welfare Council: London, UK

Farm Animal Welfare Forum 2010 *Labelling food from farm animals: Method of production labels for the European Union.* Farm Animal Welfare Forum, Godalming, UK. http://www.fawf.org.uk/ (Accessed 5 July 2011)

Green LE, Lewis K, Kimpton A and Nicol CJ 2000 Cross-sectional study of the prevalence of feather pecking in laying hens in alternative systems and its association with management and disease. *Veterinary Record 147*: 233-238. http://dx.doi.org/10.1136/vr.147.9.233

Lambton SL, Knowles TG, Yorke C and Nicol CJ 2010 The risk factors affecting the development of gentle and severe feather pecking in loose housed laying hens. *Applied Animal Behaviour Science 123*: 32-42. http://dx.doi.org/10.1016/j.applanim.2009.12.010

Main DCJ, Kent JP, Wemelsfelder F, Ofner E and Tuyttens FAM 2003 Applications for methods of on-farm welfare assessment. *Animal Welfare 12*: 523-528

Main DCJ and Mullan S 2012 Economic, education, encouragement and enforcement influences within farm assurance schemes. *Animal Welfare 21(S1)*: 107-111. http://dx.doi.org/10.7120/096272812X13345905673881

Mench JA 2003 Assessing animal welfare at the farm and group level: a United States perspective. *Animal Welfare 12*: 493-503

Mullan S, Edwards SA, Whay HR, Butterworth A and Main DCJ 2011 Inter-observer reliability testing of pig welfare outcome measures proposed for inclusion within farm assurance schemes. *The Veterinary Journal 190*: e100-e109. http://dx.doi.org/10.1016/j.tvjl.2011.01.012

Nicol CJ, Gregory NG, Knowles TG, Parkman ID and Wilkins LJ 1999 Differential effects of increased stocking density, mediated by increased flock size, on feather pecking and aggression in laying hens. *Applied Animal Behaviour Science 65*: 137-152. http://dx.doi.org/10.1016/S0168-1591(99)00057-X

RSPCA 2011 *RSPCA welfare standards for laying hens and pullets.* Royal Society for the Prevention of Cruelty to Animals: Horsham, UK

Soil Association 2011 *Soil Association organic standards for producers 2011 version 16.4 (updated June 2011).* Soil Association, Bristol, UK. www.soilassociation.org/whatwedo/organicstandards.aspx (Accessed 5th July 2011)

Webster AJF 2009 The Virtuous Bicycle: a delivery vehicle for improved farm animal welfare. *Animal Welfare 18*: 141-147

Welfare Quality® Consortium 2009a Assessment protocol for pigs. *Welfare Quality®*. NEN: The Netherlands

Welfare Quality® consortium 2009b Assessment protocol for poultry. *Welfare Quality®*. NEN: The Netherlands

Welfare Quality® consortium 2009c Assessment protocol for cattle. *Welfare Quality®*. NEN: The Netherlands

Whay HR, Main DCJ, Green LE, Heaven G, Howell H, Morgan M, Pearson A and Webster AJF 2007 Assessment of the behaviour and welfare of laying hens on free range units using animal-based measurements. *Veterinary Record 161*: 119-128. http://dx.doi.org/10.1136/vr.161.4.119

Wilkins LJ, Brown SN, Zimmerman PH, Leeb C and Nicol CJ 2004 Investigation of palpation as a method for determining the prevalence of keel and furculum damage in laying hens. *The Veterinary Record 155*: 547-549. http://dx.doi.org/10.1136/vr.155.18.547

https://doi.org/10.7120/09627286.21.3.389 Published online by Cambridge University Press