

Welfare assessment of laying hens in furnished cages and non-cage systems: assimilating expert opinion

TB Rodenburg^{*†‡}, FAM Tuytens[‡], K de Reu[§], L Herman[§], J Zoons[#] and B Sonck[‡]

[†] Animal Breeding and Genomics Centre, Wageningen University, PO Box 338, 6700 AH Wageningen, The Netherlands

[‡] Institute for Agricultural and Fisheries Research (ILVO), Animal Sciences, Animal Husbandry & Welfare, Scheldeweg 68, 9090 Melle, Belgium

[§] Institute for Agricultural and Fisheries Research (ILVO), Technology and Food Sciences, Food Safety, Brusselsesteenweg 370, 9090 Melle, Belgium

[#] Provincial Centre for Applied Poultry Research, Poel 77, 2440 Geel, Belgium

* Contact for correspondence and requests for reprints: bas.rodenburg@wur.nl

Abstract

It is extremely difficult to carry out an assessment of welfare in an entirely objective manner. The choice of welfare indicators, as well as the assignment of relative weightings to these indicators, both involve a certain degree of subjectivity. The aim of this study was to create a possible method of dealing with this subjectivity, using the opinions of groups of experts to increase the consensus for a protocol for the on-farm assessment of laying-hen welfare. The selection of the 17 separate welfare indicators was based both on a questionnaire submitted to 18 international poultry welfare experts and on the practical feasibility of collecting the respective data during a one-day farm visit. Subsequently, a second group of 13 experts was asked to assign relative weightings to the welfare indicators in this protocol. This assessment was carried out twice, once with and once without provision of detailed information on the welfare indicators. When detailed information was provided, the weightings assigned to the welfare indicators were generally lower than when no detailed information was provided. In conclusion, subjectivity regarding the choice of welfare indicators and the assignment of their relative weightings, can be dealt with and made transparent by seeking consensus among experts. Although the choice of experts, the methodology for extracting consensus data, and the nature and amount of information on the welfare indicators that should be provided, are likely to benefit from further refinement, the data presented in this study should be valuable for the development and application of formalised protocols for an integrated assessment of the welfare of laying hens, on-farm.

Keywords: animal welfare, expert opinion, furnished cages, laying hens, non-cage systems, welfare assessment

Introduction

In order to compare the welfare of animals in a range of housing systems, we require methods to perform overall welfare assessments of these systems. These assessments cannot be carried out objectively as both the choice of welfare indicators and the assignment of relative weightings to these indicators, involve a certain degree of subjectivity (Spoolder *et al* 2003). It has been shown that the choice of welfare indicators can be strongly influenced by one's view and definition of animal welfare (Fraser 2003). A possible solution would be to base both the decision on the choice of welfare indicators as well as the assignment of relative weightings to these indicators, on consensus between experts.

In recent years, a number of different welfare assessment methods have been developed and evaluated. These can use, as their basis, environment-based welfare indicators and animal-based welfare indicators, as stated by Johnsen *et al* (2001). Environment-based parameters are related to the

system, the management, and the stockmanship of the farmer, whereas animal-based parameters record the 'response' of the animals to that particular system (Johnsen *et al* 2001). For laying hens, a limited number of welfare assessment methods is available. The first method, the Animal Needs Index, is based mainly on environment-based welfare indicators, such as space allowance and housing design (Striezel *et al* 1994; Bartussek 2001). A second method, currently being developed within the European project, Welfare Quality®, is based mainly on animal-based indicators (Keeling & Veissier 2005). Other methods combine environment-based and animal-based welfare indicators (Oden *et al* 2002; Mollenhorst *et al* 2005).

Not all methods involve the integration of separate welfare indicators into an overall welfare score. It could be argued that welfare scientists, including ethologists, veterinarians and general welfare scientists, are best qualified to make a judgment of the overall welfare state, as it is very hard for

people who are not experts in the field of animal welfare to correctly interpret the complete dataset on the various welfare indicators. Animal welfare scientists should be able to assess the relevancy of certain indicators for animal welfare, although they too may require more detailed information about the precise nature of the methods used. By integrating the welfare indicators, scientists may provide a tool that can be used to compare the welfare of housing systems for laying hens and improve the welfare of laying hens on commercial farms.

There are various possible methods of integrating welfare indicators into an overall score. In the Animal Needs Index, the indicators are integrated subjectively, using weights assigned by stakeholders and the researchers that developed the method (Striezel *et al* 1994; Bartussek 2001). The advantage of this approach is that the researchers who developed the method probably have an advantage when it comes to assigning weightings to the indicators that they included in the protocol. They are familiar, presumably, with the recording method of each indicator and aware of any practical considerations (repeatability, validity, ease and accuracy of the measurement) that may be significant. An important disadvantage, however, is that the method of integrating the welfare indicators is developed by a single person or a small group of researchers, without scientific consensus among experts. Hence, the integration of welfare indicators may be fairly subjective.

By assigning weightings based on the opinion of a number of experts, the consensus for a welfare assessment protocol can be increased. Different methods can be used to assign weightings based on expert opinion. Haslam and Kestin (2003) used conjoint analysis to assign weightings. They selected six important welfare indicators with predefined possible values, ranging from poor to good welfare. The experts were presented with examples of farms, stating the value of each of the six welfare indicators per farm. Experts were then asked to give an overall welfare assessment of each farm. The individual welfare indicator weightings were extracted from the assessment using conjoint analysis. The drawbacks of this method are that only a limited number of indicators can be included and that predefined values are required. Alternatively, experts can be consulted anonymously using the Delphi-method (Anonymous 2001; Whay *et al* 2003). They are consulted in more than one consultation round. In the first round, they themselves can come up with a list of welfare issues and possible welfare indicators to address these issues. In a second round, the experts can then be asked to score the importance of each measure.

A third possibility is to assign weightings based on scientific evidence and to validate the method by comparing it with the opinion of a panel of experts. De Mol *et al* (2006) assigned weighting factors to the various welfare indicators based on scientific literature on housing, management, and welfare of laying hens, following the procedure described by Bracke *et al* (2002a, b) and created a model for laying hens containing this information. They validated this model by comparing the scores produced by the model with scores

given by experts on poultry welfare and found a large degree of accordance between the two.

There are clear advantages to assigning weightings to welfare indicators based on consensus among experts. If weightings are assigned based on consensus among experts, the subjectivity of the integration method is reduced. It is unclear, however, how much additional information on welfare indicators, experts need to come to a valid assessment.

The aim of this study was to propose a possible method of dealing with the subjectivity involved in the development of an overall welfare assessment protocol for laying hens, using expert opinion. A second aim was to evaluate the importance of detailed information on the welfare indicators for the scores of these indicators and the importance of taking practical considerations (for instance low repeatability, difficulty in measuring) into account.

Materials and methods

Selection of welfare indicators

A questionnaire was used to aid selection of welfare indicators to be included in the protocol. It was completed by 18 participants of the Seventh European Symposium on Poultry Welfare, held in Lublin, Poland from 15 to 19 June 2005. Every participant was free to fill out the questionnaire and, after the symposium, it was checked whether these individuals had previously published scientific papers on welfare of laying hens. The results of this check were such that it was not necessary to exclude any of the 18 participants. A concept-protocol to assess welfare of laying hens was presented in the questionnaire. This concept-protocol contained 19 separate welfare indicators and, for each, participants were asked to indicate the significance for animal welfare on a scale between 0 (not important at all) and 10 (extremely important). Participants were also asked to suggest additional welfare indicators to be included in the protocol. The next stage involved farm visits to study the feasibility of measuring each welfare indicator in practice. A final protocol, based on the results of the questionnaire and the farm visits was defined, which included 17 welfare indicators (Table 1). Measures included information on animal-based welfare indicators (feather condition, fear, bone fractures, mortality, behavioural observations), environment-based indicators (Animal Needs Index) and indicators related to hygiene and health (dust and bacteria levels, shell cleanliness and quality). This protocol was presented at a workshop with 13 experts, to assign weightings to each of the 17 indicators.

Assigning weightings to welfare indicators

A workshop was held at the Third International Workshop on Assessment of Animal Welfare at Farm and Group Level in Vienna, Austria on 23 September 2005. The main aim of this workshop was to assign weighting factors to the welfare indicators in the protocol. A further aim was to establish the difference between assessing the importance of welfare indicators, with or without additional information on each measurement, and with or without allowing for practical

Table 1 Proposed protocol to measure welfare of laying hens on farms. Column session 1 was the only information given to the participants for session 1. Column session 2 gives a brief description of the additional information provided for session 2.

Welfare indicator	Session 1	Session 2
Inhalable dust	Inhalable dust concentration in the air (all dust particles)	Concentration of inhalable dust measured using a personal dust sampler for 45 minutes (gravimetric analysis)
Respirable dust	Respirable dust concentration in the air (fine dust particles)	Concentration of respirable dust measured using a personal dust sampler for 45 minutes (gravimetric analysis)
Airborne bacteria	Bacteria levels in the air: total number of aerobic germs	Concentration of aerobic bacteria in the air (air sample 12.5 l)
<i>Enterobacteriaceae</i>	Bacteria levels in the air: total number of <i>Enterobacteriaceae</i>	Concentration of <i>Enterobacteriaceae</i> in the air (air sample 400 l)
Animal Needs Index	Animal Needs Index (TGI laying hens)	Animal Needs Index TGI-200 was calculated based on information provided by the farmer and on additional measurements (light intensity, litter quality)
Foraging	Litter use: amount of foraging behaviour	Mean percentage of foraging per farm (2 × 30 min behavioural observations in litter area)
Dustbathing	Litter use: amount of dustbathing behaviour	Mean percentage of dustbathing per farm (2 × 30 min behavioural observations in litter area)
Plumage condition	Feather damage	Sum of feather damage to neck, back, tail, vent, breast (for method see Tauson <i>et al</i> 2005); 30 birds per flock
Red mites	Incidence of red mites	Mean weight of red mites per trap (9 traps per farm, left overnight)
Shell cleanliness	Egg quality: percentage of dirty/cracked eggs	Visual assessment of cleanliness of 60 eggs per farm
Shell quality	Egg quality: eggshell breaking strength	Assessment of shell breaking strength of 60 eggs per farm
Shell bacteria	Egg quality: bacteria on the eggshell	Assessment of bacterial load of 60 eggs per farm
Perch use	Behavioural observations: perch use	Mean percentage of daytime perch use per farm (2 × 30 min behavioural observations in perching area)
Bone fractures	Bone breaks to the keel bone and furculum	Percentage of birds with keel bone fractures per flock (dissection of 15 birds per flock)
Bone strength	Bone strength of the leg and wing bones	Mean bone strength of leg, wing and keel bones (dissection of 15 birds per flock)
Fear	Fearfulness: tonic immobility	Tonic immobility test, where bird was restrained on the back. Mean latency to stand up (15 birds per flock)
Mortality	Mortality	Mortality levels up to 60 weeks of age from farmer's records

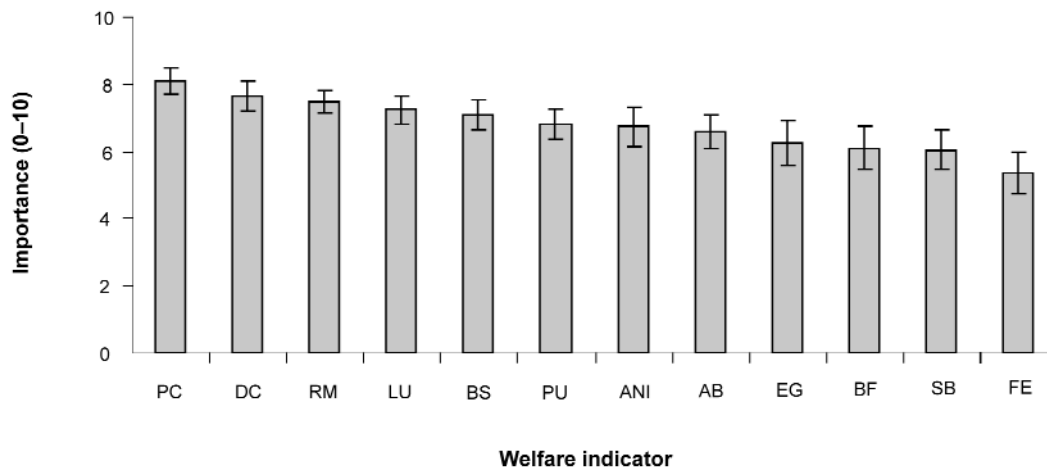
considerations (for instance low repeatability, difficulty in measuring). Overall, the workshop sought to increase consensus for our welfare assessment protocol by assigning weighting factors to welfare indicators for laying hens based on the opinion of an expert panel. Firstly, the (13) participants were presented with the list of 17 welfare indicators and asked to indicate the importance of each measurement to the welfare of laying hens on a scale from 0 to 10. The only information provided at that time can be seen in Table 1 (column session 1). They were asked to assess the theoretical validity of the indicator, without taking practical considerations into account. Following on from this, a short presentation was given on each measurement, including the relevance to welfare, the method of measuring and the actual measurement data. A summary of the information provided is given in Table 1 (column session 2). Also, room for questions and discussion was provided. Subsequently, the

participants were asked to assess the importance of the indicators for a second time, without access to their original results. In this session, they were asked to take into account additional information on the indicators, as well as any practical considerations (for instance, low repeatability, difficulty in measuring) into account.

Statistical analysis

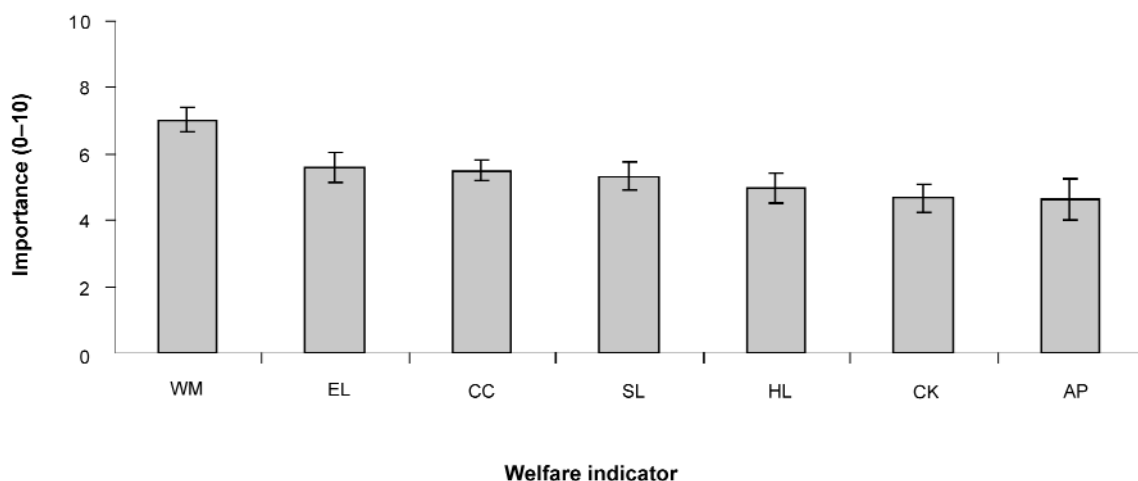
The experts' scores from sessions 1 and 2 were analysed in SAS using repeated measures analysis of variance. The scores for each of the 17 separate welfare indicators were analysed as dependent variables using the GLM procedure with session and expert included as independent class effects. To calculate agreement between experts, Pearson correlations were calculated between the scores of each pair of experts. The mean of all these correlations was used as a measure of agreement between experts, for sessions 1 and 2, separately.

Figure 1



The importance of the welfare indicators that were finally included in the protocol, according to the participants of the questionnaire (0 is not important, 10 very important). PC (plumage condition), DC (dust concentration), RM (red mites), LU (litter use), BS (bone strength), PU (perch use), ANI (Animal Needs Index), AB (airborne bacteria), EG (egg quality), BF (bone fractures), SB (shell bacteria), FE (fear).

Figure 2



The importance of the welfare indicators that were finally excluded from the protocol, according to the participants of the questionnaire (0 is not important, 10 very important). WM (worms), EL (egg location), CC (corticosterone), SL (*Salmonella*), HL (heterophil/lymphocyte ratio), CK (creatin kinase), AP (acute phase protein).

Results

Selection of welfare indicators

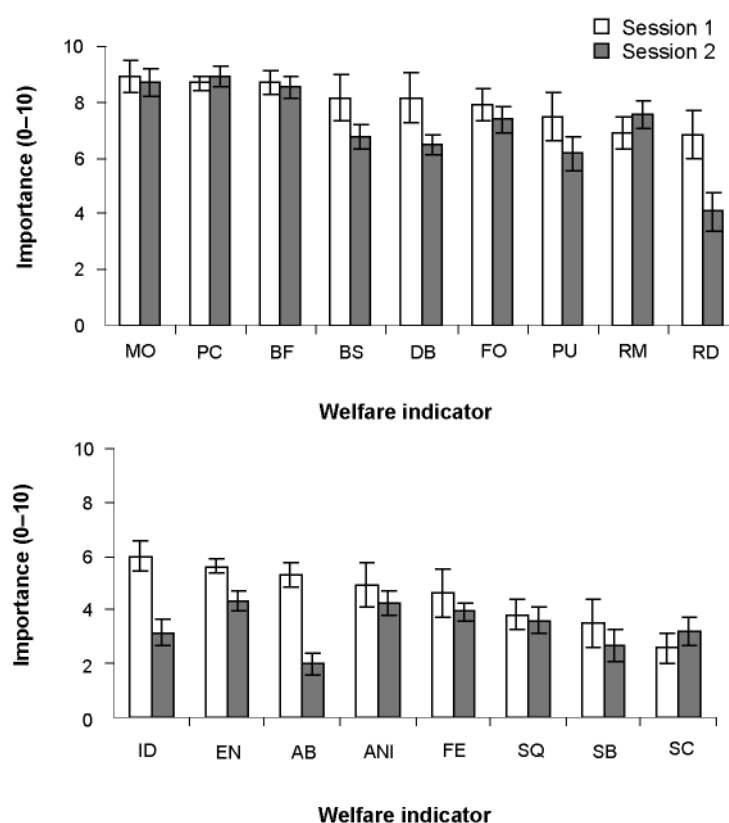
The ranking of the welfare indicators is shown in Figures 1 and 2. Based on the assessment by the participants to the questionnaire and based on practical considerations from the farm visits, 12 welfare indicators were retained in the protocol (see Figure 1) and seven were excluded (see Figure 2).

The practical considerations for excluding certain welfare indicators were as follows: i) it was not possible to collect litter

samples from furnished cages therefore the presence of worm eggs and corticosterone in the litter could not be studied; ii) nest use was difficult to record in furnished cages because, on some farms, the egg belt continued to move throughout the day, making it impossible to study where the eggs were laid and iii) in our study, it was not feasible to collect blood samples, on-farm. The indicators for which blood samples were required also did not generally receive high scores from the experts therefore the anti-titre against *Salmonella* spp, the heterophil/lymphocyte ratio and the concentration of acute phase proteins were excluded from the protocol.

Figure 3

The importance of the nine highest scoring (upper) and the nine lowest scoring (lower) welfare indicators in the sessions without (1) and with (2) additional information about the indicators (0 is not important, 10 very important). Upper: MO (mortality), PC (plumage condition), BF (bone fractures), BS (bone strength), DB (dustbathing), FO (foraging), PU (perch use), RM (red mites), RD (respirable dust); Lower: ID (inhalable dust), EN (*Enterobacteriaceae*), AB (airborne bacteria), ANI (Animal Needs Index), FE (fear), SQ (shell quality), SB (shell bacteria), SC (shell cleanliness).



Following suggestions by participants to the workshop and the first farm visits, four of the welfare indicators were split into two separate categories: dust was split into either inhalable or respirable dust; litter use was split into foraging and dustbathing; airborne bacteria was split into total airborne bacteria and *Enterobacteriaceae* spp and egg quality was split into cleanliness of the eggshell and bacteria on the eggshell (see Table 2). Furthermore, mortality was added to the list, resulting in a final protocol consisting of 17 welfare indicators which was presented at a workshop with 13 experts for assignment of weightings to each of the indicators.

Assigning weightings to welfare indicators

In the workshop, participants were asked to assign weightings in two separate sessions. In session 1, no additional information on welfare indicators was given, whilst in session 2, the participants were given additional information and asked to take practical considerations (for instance, low repeatability, difficulty in measuring) into account. In session 1, mean scores ranged from 2.58 for shell cleanliness to 8.92 for mortality (Figure 3). In session 2, mean scores ranged from 2.00 for airborne bacteria to 8.92 for plumage condition. In session 1, scores were generally higher than in session 2. In particular, the concentrations of respirable dust ($F_{1,24} = 8.67$; $P < 0.01$) and inhalable dust ($F_{1,24} = 6.62$; $P < 0.05$) and the total number of airborne bacteria

Table 2 The welfare indicators that were split into two separate welfare indicators in the final protocol.

Concept protocol	Final protocol
Dust	Respirable dust Inhalable dust
Litter use	Foraging Dustbathing
Airborne bacteria	Total airborne bacteria <i>Enterobacteriaceae</i>
Egg quality	Bacteria eggshell Cleanliness eggshell

($F_{1,24} = 10.89$; $P < 0.01$) were given a lower assessment in session 2 compared to session 1. Mean correlations between expert scores were 0.49 for session 1 and 0.47 for session 2.

Discussion

Welfare indicators

The aim of this study was to propose a possible method of dealing with the inherent subjectivity involved in developing an overall welfare assessment protocol for laying hens, using expert opinion. Based on the assessment provided by participants in the questionnaire and applica-

tion of the protocol, on-farm, 12 of the original set of welfare indicators were retained in the welfare assessment protocol and seven excluded. In addition, some new indicators were included, such that the final protocol consisted of 17 indicators that could be measured in a comparable way in all housing systems for laying hens. This protocol was then presented to participants in the workshop in order to assign weighting factors to each separate indicator.

In session 1 of the workshop, during which the participants had no additional information on the indicators and were asked not to take practical considerations (for instance low repeatability, difficulty in measuring) into account, scores were generally higher than in session 2. In particular, the concentrations of respirable and inhalable dust and the total numbers of airborne bacteria were given lower weightings in session 2 compared to session 1. These lower weightings for session 2 could be due to the additional information, to the taking into account of practical considerations, to the discussion during the workshop or to a combination of all these factors.

Regarding dust concentration, a participant in the workshop stated that dust does not represent a major problem for laying hens as their respiratory system differs from that of mammals and is less sensitive to dust. This could be a major factor why we saw lower scores assigned to inhalable and respirable dust in session 2. Guarino *et al* (1999), however, did find that a higher dust concentration in a layer house was correlated with increased mortality. Regarding airborne bacteria, again, it may have been the case that the discussion during the workshop influenced the results. A participant stated that total numbers of airborne bacteria was not a good measure of animal welfare. This illustrates a potential problem of our methodology: on the one hand, the expertise and knowledge of the experts is useful in helping to further improve the welfare assessment system, on the other, though, participants can make certain statements with great authority, without organisers being able to check the scientific evidence behind that statement. Agreement between experts was moderate for session 1 ($r = 0.49$) and 2 ($r = 0.47$). There was a degree of variation in scores between experts in both sessions. We had expected there to be greater agreement for session 2 compared to session 1, as all the experts had been present at the discussion on welfare indicators, but this did not prove to be the case.

Experts were asked directly in this study to assign weightings to separate welfare indicators. Other methods, such as conjoint analysis (Haslam & Kestin 2003) involve more indirect methods to assign weightings. In their study, experts were presented with examples of farms, stating the value of each of the six welfare indicators per farm. Experts were then asked to give an overall welfare assessment of each farm.

The individual welfare indicator weightings were extracted from the assessment using conjoint analysis. The advantage of using a direct method, in the present study, was that experts could consider indicators and, likewise, additional information and practical considerations could also be looked at separately. A disadvantage of this direct method is that it appears to invoke greater levels of resistance, as scientists are hesitant to directly assign weight-

ings to welfare indicators. At the workshop in Vienna, we agreed that it is possible to combine welfare indicators that are all related to physical health (mortality, bone breaks, disease incidence) into a single score. The same rule would apply to welfare indicators that are related to mental health (space per bird, ability to perform natural behaviour). The general feeling of the majority of scientists present at the workshop was that integration of the physical and mental 'score' should be left to politicians. It could be argued, however, that politicians would again need scientific opinion to come to this integration.

When integrating separate welfare indicators into an overall welfare score, it is important to consider the degree to which compensation of welfare problems should be allowed, ie the notion that a serious welfare problem can be compensated for by a number of minor advantages (Spooler *et al* 2003). Bracke *et al* (2002b) introduced 'vetoes', which are very high weighting factors (for instance 1,000 compared to 2 or 3) that dictate minimum scores for certain welfare indicators. It would be possible to ask experts to indicate which scores would qualify as vetoes and compensation could be limited by allowing integration only within the various welfare dimensions (ie, to the level of physical and mental health or the five freedoms) and not between dimensions, as was also suggested by some participants to the workshop.

An important conclusion reached by the workshop is that the provision of additional information about welfare indicators and discussion regarding parameters can affect the evaluation of the parameters. The question is how much information should an expert have in order to come to a good assessment which can be used to assign weighting factors to the different welfare indicators. Our impression, from the workshop, is that it is useful to provide detailed information on how each parameter is measured. It is a matter for debate whether it is a good idea to allow discussion about parameters prior to scoring; in a workshop there is always the risk that some participants will be influenced by the opinion of other, more dominant, participants. Furthermore, it is difficult to check whether statements made during the workshop are true and to what extent they are supported by the literature. The scores from the session which included additional information may still be of more use when calculating an overall welfare score, as it is important for such an overall score that the individual welfare indicators are measured precisely and that the indicators are relevant to animal welfare. These weighting factors are essential for aggregating the scores of single welfare indicators into an overall welfare assessment.

Conclusions and animal welfare implications

The subjectivity regarding the choice of welfare indicators and the assignment of their relative weightings can be dealt with and made transparent by seeking consensus among experts. Although the choice of experts, the methodology for extracting consensus data and the nature and amount of information on the welfare indicators that should be provided are likely to benefit from further refinement, the data presented in this study should be valuable for the devel-

opment and application of formalised protocols for an integrated assessment of the welfare of laying hens, on-farm.

Acknowledgements

The authors would like to thank all who participated in the questionnaire and to the workshop for their invaluable contributions and suggestions. Furthermore, we would like to thank the organisers of the Seventh European Symposium on Poultry Welfare in Lublin, Poland and the organisers of the Third International Workshop on Assessment of Animal Welfare at Farm and Group Level in Vienna, Austria for giving us the opportunity to present our questionnaire and to organise the workshop, respectively. This project was funded by the Department of Health, Food Chain Safety and Environment of the Federal Government of Belgium.

References

- Anonymous** 2001 Scientists' assessment of the impact of housing and management on animal welfare. *Journal of Applied Animal Welfare Science* 4: 3-52
- Bartussek H** 2001 An historical account of the development of the animal needs index ANI-35L as part of the attempt to promote and regulate farm animal welfare in Austria: an example of the interaction between animal welfare science and society. *Acta Agricultura Scandinavica, Section A, Animal Sciences* 30: 34-41
- Bracke MBM, Metz JHM, Spruijt BM and Schouten WGP** 2002a Decision support system for overall welfare assessment in pregnant sows B: Validation by expert opinion. *Journal of Animal Science* 80: 1835-1845
- Bracke MBM, Spruijt BM, Metz JHM and Schouten WGP** 2002b Decision support system for overall welfare assessment in pregnant sows A: Model structure and weighting procedure. *Journal of Animal Science* 80: 1819-1834
- de Mol RM, Schouten WGP, Evers E, Houwers HWJ and Smits AC** 2006 A computer model for welfare assessment of poultry production systems for laying hens. *Wageningen Journal of Animal Sciences* 54: 157-168
- Fraser D** 2003 Assessing animal welfare at the farm and group level: the interplay of science and values. *Animal Welfare* 12: 433-443
- Guarino M, Caroli A and Navarotto P** 1999 Dust concentration and mortality distribution in an enclosed laying house. *Transactions of the ASAE* 42: 1127-1133
- Haslam SM and Kestin SC** 2003 Use of conjoint analysis to weight welfare assessment measures for broiler chickens in UK husbandry systems. *Animal Welfare* 12: 669-675
- Johnsen PF, Johannesson T and Sandøe P** 2001 Assessment of farm animal welfare at herd level: many goals, many methods. *Acta Agricultura Scandinavica, Section A, Animal Sciences* 30: 26-33
- Keeling LJ and Veissier I** 2005 Developing a monitoring system to assess welfare quality in cattle, pigs and chickens. In: Butterworth A (ed) *Science and Society Improving Animal Welfare, Welfare Conference Proceedings*. 2005, Brussels, Belgium
- Mollenhorst H, Rodenburg TB, Bokkers EAM, Koene P and de Boer IJM** 2005 On-farm assessment of laying hen welfare: a comparison of one environment-based and two animal-based methods. *Applied Animal Behavioural Science* 90: 277-291
- Oden K, Keeling LJ and Algers B** 2002 Behaviour of laying hens in two types of aviary systems on 25 commercial farms in Sweden. *British Poultry Science* 43: 169-181
- Spolder H, de Rosa G, Horning B, Waiblinger S and Wemelsfelder F** 2003 Integrating parameters to assess on-farm welfare. *Animal Welfare* 12: 529-534
- Striezel A, Andersson R and Horning B** 1994 Animal Needs Index for laying hens. In: Sundrum A, Andersson R and Postler G (eds) *Animal Needs Index 200: a Manual to Assess Housing Systems* pp 73-112. Kollen Druck & Verlag: Bonn, Germany
- 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