University of Reading

Valuation of animal welfare benefits

A report to Defra by

Richard Bennett[1], Andrew Butterworth[2], Philip Jones[1],

Ariane Kehlbacher[1] and Richard Tranter[1]

March 2012

[1] School of Agriculture, Policy and Development
University of Reading
Earley Gate
Reading
RG6 6AR

[2] Animal Behaviour and Welfare Group
School of Veterinary Sciences
University of Bristol
Langford House
Langford
Bristol
BS40 5DU

Contents

Executive summary

Objective and aims

1. While the costs of delivering improvements in animal welfare are relatively straightforward to measure, the benefits that these bring are not well understood. One consequence is that it is often not possible to estimate the true value of investment in welfare improvements. The objective of this study is the enhancement of our ability to both measure welfare improvements and value the benefits delivered by actions intended to improve the welfare of different kept animals, especially farmed animals.

2. The aims of the project were to:

(i) carry out a comprehensive literature review relevant to both animal welfare science and the economic valuation of benefits;
(ii) carry out an assessment of the total potential consumer spend on animal welfare, including charitable giving; the development of a theoretical framework for the valuation of animal welfare benefits; and
(iii) critically appraise the evaluation framework that was designed, clearly identifying the merits and limitations of the methodology.

Project organisation

3. The project started on 15 June 2010 and a draft final report was submitted on 25 February 2011. The final report was submitted on 9 March 2012.

4. The project team consisted of a veterinarian, three agricultural economists and a psychologist. The team, with research experience totalling over 130 years, was led by Professor Richard Bennett, a member of FAWC. A workshop containing the project Steering Group who were mainly animal welfare scientists and specialists in economic valuation techniques, as well as some Defra officials, helped the research team clarify their ideas mid-way through the project.

Welfare assessment of animals

5. Existing animal welfare assessment schemes on farms, in laboratories or in zoos, tend to assess welfare by examination of the provision of housing or resources (resource based measures, RBMs), rather than looking at outcomes for the animals (outcome based measures, OBMs). Scientists have suggested that OBMs can provide better indicators of animal welfare, as welfare is a characteristic of the individual animal, not just of the system in which they are kept. It is argued that the sorts of questions that need to be asked are:

(i) are the animals properly fed and supplied with water?
(ii) are the animals properly housed?
(iii) are the animals healthy?

(iv) can the animals express a range of behaviours and emotional states?

Many people feel that a balanced mixture of animal and resource based measures are likely to be the way in which welfare measurement may move forward in the future.

6. Many farms, zoos, aquaria and other facilities are keen to identify welfare monitoring tools that can be readily applied and allow for rapid assessment to provide timely feedback for management. To implement effective use of animal based measures, it is necessary to adopt a number of logical steps. First, measure the 'level' of welfare condition (using both RBMs and OBMs), then analyse the risk factors - the variables which may influence the development of welfare 'issues' - before informing the producer, animal keeper or zoo (and perhaps the user or purchaser) what has been found. Finally, they should support management decisions necessary to create improvements in welfare.

7. The use of OBMs allows the creation of a range of 'scores' which can be combined and presented to the producer (animal carer, zoo etc.) and even to consumers. The Welfare Quality tool already has the capacity to give an aggregated score to a farm through use of a webtool based calculation method which returns a score if farm data is input. This requires the attribution of weighted values to the measures, to assess the impact of each measure with respect to animal welfare. A multi-pronged strategy involving various RBMs and OBMs is most likely to provide the capacity for overall welfare monitoring of animal welfare in the future. Such a strategy may include regularly updated species care guidelines, accreditation standards, longitudinal studies and the development of practical assessment tools (including potentially automated systems using cameras) for welfare monitoring.

Methods for valuing farm animal welfare

8. The different approaches that can be used to value private animal welfare benefits (PAWB), which are the utility that consumers of high animal welfare status products derive from the perceived better taste, food safety, healthiness or nutritional value, and social animal welfare benefits (SAWB), which are the benefits that are experienced by those who care about the livestock production process in relation to environmental, ethical, or other reasons are reviewed and appraised. Stated preference methods that have been widely used to value farm animal welfare include the contingent valuation method and the choice experiment method. Both methods can value PAWB and SAWB and can be used for *ex-ante* analysis of policy measures. The choice experiment method is also able to value species trade-offs.

9. Valuation methods that can only value PAWB, and have been applied to animal welfare in the past, include non-hypothetical choice experiments, auction based methods and hedonic pricing with the latter only being able to measure the product characteristics of which individual consumers are aware. Other valuation methods that have not yet been used to value farm animal welfare and are probably inherently less suitable to do so, include the travel cost method, averting expenditure and subjective well-being methods.

Consumer spend on animal welfare

10. It was found that, in 2009, around £1,700 million was spent in the UK by consumers on animal welfare in its broadest sense. This was made up of £175 million charitable spend associated with animal welfare and the rest for five broad categories of brands and products with animal welfare attributes. However, whilst most of the latter category went on farmed animal welfare, probably only 10% of the charitable spend associated with animal welfare was on farmed animal welfare.

The policy evaluation tool

11. A tool that would allow policy makers to undertake either *ex-ante* or *ex-post* evaluations of policy proposals that impact on the keeping of farmed animals was designed. This tool projected both the extent of changes to animal welfare and also attributed valuations to these changes. The functional requirements of the tool were discussed across a number of Defra requirements. A Multiple Criteria Analysis (MCA) of existing welfare measurement approaches was first undertaken to examine the strengths and weaknesses of each. Three extant welfare measurement approaches were evaluated - the German Animal Needs Index 200, the Austrian ANI 34L systems, and the multi-country Welfare Quality Protocol (WQ). Complete welfare measurement systems were chosen as they already encapsulated the extremely large and very disparate body of scientific and practitioner data on the drivers and measurement of (farm) animal welfare.

12. The MCA exercise revealed that none of the existing welfare measurement systems were wholly suitable for use as the basis of the new tool, although the closest in terms of functionality was the WQ approach. This approach was, therefore, adopted as a basis for the new welfare measurement tool, subject to some extension and modification. It is proposed that the new welfare measurement tool captures all links in the chain of causality between policy and welfare state i.e. policy —> environment state —> health & mental state —> welfare, while the WQ approach captures only the latter two of these stages. Understanding the link between policy and the environment is relatively straightforward because the conceptual gulf is small and because policies are usually framed in terms of changes to the environment and management practice. New environment settings arising from policy proposals would then be analysed in terms of their impact on the health and mental states of animals.

13. Modelling this environment to health link in the chain of causality is more difficult, as single aspects of the environment are known to impact on multiple dimensions of physical and mental animal state. Using an expert group, this link in the chain can be created by first mapping the key relationships between aspects of the environment and dimensions of physical state, then attaching weights to each relationship to reflect the importance of each environment driver. In the modelling process, new environmental states arising from policy changes can be scored by an expert group and these scores weighted and summed to arrive at new settings for a range of measures of health and mental state. Alternatively, where appropriate animal welfare impact assessment trial or field data are available, these can be

used to populate the required values of welfare parameters (health and mental state scores) within the tool.

14. In the next link in the chain, the health and mental state scores would be weighted and combined using a process very similar to that employed in the WQ protocol, to create a single aggregate welfare score i.e. the welfare measure. Some adaptation of the WQ approach would be required to correct for methodological inconsistencies with the tool requirements. An expert group would be needed to generate these coefficients at the model construction stage.

15. Once the model is constructed, its operation requires a process involving limited use of expert groups. This process would involve three stages: application; validation; and updating and calibration. In the application phase expert input would be required to interpret new policy proposals in terms of changes to environment conditions. The model operation would be automated thereafter. The validation phase would occur only infrequently as data on the historic performance of the model becomes available, especially through farm surveys. One of the strengths of this modelling approach is that because data are generated on the health and welfare states of animals based on a range of detailed indicators, the welfare of animals can be surveyed on farm using these very same indicators, allowing for very detailed assessment of model performance. The updating stage would also occur only infrequently, with the coefficients embedded in the model being revisited in light of new developments in scientific understanding.

Economic valuation of improvements in animal welfare

16. Policy makers need information about the extent and nature of the benefits that arise to society as a result of improved farm animal welfare, for example, for use in a cost-benefit assessment of policy. This information is needed over a range of welfare levels because, with changing farming techniques and individual preferences, people's benefit estimates will change in the future. The aim of this part of the research was to explore a method to value different levels of improvements in beef cattle, pig and broiler chicken welfare, with particular consideration of the benefits to consumers from consuming high welfare status meat and the benefits to society, in general, from introducing measures to improve farm animal welfare.

17. For this purpose, an animal welfare score was developed and a choice experiment study and also a contingent valuation study were carried out in Great Britain. The choice experiment data were analysed using the random parameter logit model, and the contingent valuation data were analysed using the ordered probit and uncertainty interval regression models. In the estimation for both studies, Bayesian inference was used.

18. The findings of this study were that the valuation tool appeared to work well. Citizens in the survey samples understood and accepted the concept of the welfare score and provided consistent and apparently rational responses and values. The estimated WTP values for welfare improvements were credible in comparison with the price premia paid for livestock

products with perceived welfare attributes. Moreover, separate surveys applying CE and CV WTP elicitation and estimation methods produced very similar estimates of people's WTP to see an improvement in welfare score.

Conclusions and recommendations

19. The Welfare Quality framework for measuring the welfare of farm animals is not entirely suitable for use in policy evaluation, especially desk-based *ex ante* evaluations. However, with appropriate modification, the Welfare Quality framework can be used to assess the welfare implications of policy changes expressed as a welfare metric, which can then be used to obtain societal valuations associated with the welfare benefits accruing from policy changes. For this to happen, it is considered necessary that the welfare status of farm animals be communicated to respondents in valuation surveys in the form of a single score on a continuous scale. In this way, marginal valuation of changes in the level of the score (which represents changes in the level of welfare) can be obtained for different farmed species.

20. Various economic valuation techniques can be applied to estimate people's willingness to pay for improvements to welfare/welfare scores, such as contingent valuation and choice experiment methods. Clearly, the results of applying these techniques within a framework such as that presented in this study, need to be treated with some circumspection. However, they can provide reasonably reliable estimates of WTP providing the survey instruments used are carefully designed and tested and providing that results are demonstrably replicable and consistent with one another, as well as with available revealed preference (i.e. market) data.

21. The review of literature revealed that there is no body of existing valuation data available that could be used to provide valuations of any welfare changes projected by the policy evaluation tool. For this reason, it would be necessary to elicit these valuations, perhaps building upon those elicited in a recent study.

22. A further national WTP elicitation study, which could follow the same format as the study referenced here, is needed to:

(i) replicate and validate the data obtained in the existing study (i.e. for dairy, pigs and poultry); and
(ii) expand the data coverage to other farmed species.

Periodic replication of this type of study would reveal any trends within society leading to any altered valuations.

23. The following actions are recommended:

(i) The societal valuations of welfare scores can be used with a range of welfare measurement systems as long as these systems generate a single welfare measure on a cardinal, or ordinal, scale. Therefore, it is thought timely to undertake a large-scale national willingness to pay study to derive societal valuations of changes to welfare scores for a range of farmed species. These outputs should then be compared with the outputs of other studies for the purposes of validation.

(ii) A research study should be commissioned by Defra to pilot the methodology that has been developed for the policy evaluation tool in the current study. This would test the technical and practical feasibility of applying the tool. This pilot need not construct a full-scale version of the tool but it should be sufficient to identify the practical merits and limitations of the tool.

(iii) There needs to be periodic collection of data from on-farm welfare assessments (e.g. using the Welfare Quality or adapted protocols) for the purposes of:

- deriving inputs into the policy evaluation tool; and
- informing expert stakeholders for the purpose of updating and revising the technical coefficients in the tool.

1. Introduction

There has been increasing interest in, and concern over, the welfare of farm animals in the EU in recent decades, particularly in the UK, which to some extent has taken the lead on animal welfare issues within Europe. Various regulatory tools have been used to protect and promote animal welfare most notably, in the UK, the Animal Welfare Act 2006 and the secondary legislation associated with this and various welfare codes of practice. However, regulation and other means to improve animal welfare incur industry costs and there is a need to balance these costs against the benefits gained.

Building on McInerney (1994), Bennett (1995) presents a theoretical framework for the valuation of animal welfare and discusses the need for economic assessments of the benefits perceived to accrue from initiatives to improve animal welfare, for example by using methods such as contingent valuation (CV). The first such attempt to value farm animal welfare benefits was by Bennett and Larson (1996) who used CV to elicit people's willingness to pay to support legislation to ban the use of battery cages for laying hens. Since then there have been a number of studies involving some element of valuation of perceived animal welfare benefits associated with policy changes or changes in practice, using a variety of valuation techniques, but primarily CV and choice experiments (CE).

However, one of the limitations of the practical application of early theoretical frameworks to value animal welfare has been a lack of understanding of the technical relationship between animal production practices and associated states of animal welfare. This problem largely stemmed from an inability to actually measure animal welfare. McInerney (2004) encapsulated the problem by saying 'If there is a willingness to pay specifically for (animal) welfare enhancements, there must exist some notion of a demand curve for animal welfare … But what does the quantity axis show? … What does 'more' animal welfare mean?' The issue of the measurement of animal welfare has been the biggest constraint to the valuation of animal welfare benefits. This lack of a robust, information-based methodology for measuring animal welfare has led some researchers to trial alternative 'expert' systems to measure welfare. For example, Bennett (*et al.* 2004) used large a Delphi panel of experts to evaluate animal welfare benefits using a weighted scoring system where the weights were determined by the expert panel.

However, recent advances in both animal welfare science, and in the practicalities of measuring the welfare status of animals, by a combination of resource/input and welfare outcome criteria, have provided the potential to overcome this constraint. The recently completed EU-funded FP6 Welfare Quality project, involving a host of institutions and researchers throughout the EU, has helped to synthesise this research and develop practical protocols for measuring animal welfare for some farm animal species. This means that, potentially, there are new possibilities available to economists to value animal welfare benefits. Indeed, recent research at Reading by Kehlbacher (2010) explored the use of an animal welfare outcomes assessment protocol to provide a credible means for measuring the welfare status of farm animals on a continuous welfare scale. This assessment protocol was

then tested by presenting movements along the scale to a sample of the GB population and different elicitation methods (CE or CV) were used to elicit their willingness to pay for quantified welfare improvements for the main livestock species using different payment vehicles (product price or taxation).

Market data currently available provide some information on the price premia that some consumers pay for animal products that have animal welfare attributes. However, the markets for such products are still not well developed and the animal welfare attributes of products are not always clear to consumers. Moreover, market data do not capture the negative externality/public good aspects of animal suffering and animal welfare, nor does market behaviour capture the true value for welfare held by citizens (see Mayfield *et al.*, 2007). While it is certainly true that the money values of spending on products with welfare attributes do not equate with economic values (McInerney, 2004), market information can provide some useful data on consumer preferences for welfare associated with market goods.

Clearly, there are a number of potential variants (e.g. combined methods) of a framework for valuing animal welfare benefits. It is the primary purpose of the research reported here to review and evaluate these methods and then to propose a cost-effective and practical framework for valuing animal welfare benefits associated with changes in policy or practice.

A key aim of the Government's Animal Health and Welfare Strategy is to ensure 'a clearer understanding of costs and benefits' associated with Government intervention. In addition, an important goal of Defra's Animal Welfare Delivery Strategy is to ensure that animal welfare policy is based on sound scientific research, practical experience and other relevant evidence. The research reported here is in pursuit of these twin objectives and has as its main aim the development of a scientific approach (incorporating both animal/veterinary science and social science) that can help to support policy decisions regarding animal welfare and make the costs and benefits of policy more transparent.

2. Objectives and aims

While the costs of delivering improvements in animal welfare are relatively straightforward to measure, the benefits that these welfare improvements bring are not well understood. One consequence of this is that it is often not possible to estimate the true value of investment in welfare improvements, or conduct cost benefit analyses of proposed policies to change levels of animal welfare. The broad objective of the study reported here is the enhancement of our ability to value the various benefits delivered by actions intended to improve the welfare of animals, taking into account the welfare status of the animals concerned, for different types of kept animal, especially farmed animals.

The technical and scientific aims of the project were to:

(i) Carry out a comprehensive review of academic and other literature relevant to both animal welfare science and the economic valuation of benefits;

(ii) Carry out an assessment of the total potential consumer spend on animal welfare, including charitable giving, plus an exploration of the validity and feasibility of trying to estimate this aggregate measure;

(iii) The development of a theoretical and practical framework for the valuation of animal welfare benefits; and

(iv) Critically appraise the evaluation framework that was designed, clearly identifying the merits and limitations of the methodology. In addition, a review of the validity and completeness of the data used to populate the framework, together with an assessment of additional data requirements was also to be carried out.

3. Method, timing of research and report structure

The project started on 15 June 2010 and was finished by 25 February 2011.

In the first few days of the project an inception meeting was held at the offices of Defra in London at which representatives of Defra made clear to the research team exactly what was required of the project (Milestone 1).

A project Steering Group consisting of animal welfare scientists and specialists in economic valuation techniques was set-up and a workshop including its members took place in Reading in October 2010 to discuss progress with the research team and to 'steer' their progress towards the design of a workable framework for the valuation of animal welfare benefits (Milestone 2). Discussion at the workshop centred round past work that had looked at the public's willingness to pay for welfare friendly produced food.

The initial stage of the research programme consisted of a comprehensive review of both British and international literature (Milestone 3). This had two strands to it. First, a review of animal welfare and livestock production systems concentrating on welfare assessment and animal welfare measures such as welfare outcome measures. The findings of the recently completed EU-funded Welfare Quality project receive especial attention in this review. This is reported briefly in Section 4 of this report with additional information and examples being given in Appendix A. Second, a review of the application of economic valuation to animal welfare from both the theoretical and applied perspectives. This included a meta-analysis of over 50 animal welfare valuation studies that have attempted to value animal welfare using either stated preference methods, hedonic pricing and travel cost and averting expenditure methods. This is reported in Section 5 of this report.

Next, an assessment of the total potential consumer spend on animal welfare was carried out (Milestone 4). This assessment, which is reported on in Appendix B, involved constructing an estimate of the market size of brands and products with animal welfare attributes as well as reviewing charitable giving associated with animal welfare.

A theoretical and practical framework for the valuation of animal welfare benefits (Milestone 6) was developed, building on the above reviews. This can be seen as a policy evaluation tool and is reported on in Section 6. The framework is broad in both content and scope and incorporates the very latest understandings of welfare standards and benefits.

Section 7 reports on the theoretical and practical development of a framework for the valuation of animal welfare benefits as potentially measured using the welfare measurement metric (or tool) expounded in Section 6.

Finally, for Milestones 7 and 8, and reported in Section 8, a critical appraisal of the full evaluation framework was carried out including a review of the validity and completeness of the data used to populate the framework.

4. Welfare assessment of animals

4.1 Introduction

European consumers expect their animal-related food products to be produced and processed with respect for the welfare of the animals (Veissier *et al.*, 2008). In a recent EuroBarometer study (http://ec.europa.eu/food/animal/welfare/euro_barometer25_en.pdf), 54% of respondents said that they had difficulties in finding adequate information on the animal welfare standards applied in producing food. The European Commission Community Action Plan on the Protection and Welfare of Animals 2006-2010 outlines a range of co-ordinated actions in this area (http://ec.europa.eu/food/animal/welfare/actionplan). Welfare outcome measures are scientific assessments carried out on animals to help understand how they respond to how they are being kept. Welfare outcomes - sometimes called 'animal based measures' (ABMs) or 'outcome based measures' (OBMs) - usually involve direct assessment of the animal itself, through observation of behaviours, through measurement of pathologies such as lameness or skin lesions, or through interpretation of records which tell you about the health status of the animal - e.g. mastitis records, treatment records, or the number of animals which have had to be humanely destroyed during an experiment (laboratory animals), or during the production cycle (farm animals). ABMs contrast with 'resource based measures' (RBMs), which rely on assessment of the provision of, for example, space, light (day and night patterns for example), bedding, feed, access to other animals of the same species, or access to zoned areas within a cage, or in a zoo or farm environment. Welfare outcome measures are usually applied *'in situ'* i.e. they are carried out where the animals are kept or housed, and they reflect what the animal experience actually IS rather than what it SHOULD be, given a particular set of resources. By assessing the animals rather than the resources, it is possible to determine how the animal responds to the resources which are provided in a lab, zoo or a farm.

As regards zoo animals, the 'viewing public', in many parts of the world, no longer view zoos as menageries, but now have an expectation to see animals in naturalistic enclosures and for the animals to able to carry out what they (the public) consider to be natural behaviours. Zoo animal assessment is starting to develop as a way of assessing the responses of captive zoo animals to the resources provided for them and, in a similar way, in the laboratory, the responses of the animals to the resources provided in their kept environment, as well as the animals' responses to the procedures which are carried out on them. For farmed animals, the consumer influences the way that animals are viewed and, particularly in Europe, consumers expect their animal-related food products to be produced and processed with respect for the welfare of the animals and the focus of EU agricultural policy is increasingly on quality rather than quantity. Traditional price mechanisms do not always allow for considerations like animal welfare to be properly recognised in the prices paid to producers and, in general, compulsory labelling of food with animal welfare information has not been adopted. Existing assurance schemes tend to assess welfare by examination of the provision of housing or resources rather than looking at the animals themselves. Practically speaking, it may be possible to combine RBMs and ABMs. For example, if animals are lame (assessed using an ABM) it may have been possible to predict lameness if the floor condition is poor (a RBM).

Research scientists have for some time suggested that welfare outcome measures could provide valid indicators of animal welfare, since welfare is a characteristic of the individual animal, not just of the system in which animals are farmed, kept captive or held for research purposes. The sorts of questions which are now being asked are:

- Are the animals properly fed and supplied with water?
- Are the animals properly housed?
- Are the animals healthy?
- Can the animals express a range of behaviours and emotional states?

4.2 Laboratory animals

To comprehensively evaluate health and welfare in laboratory animals requires a holistic assessment that includes not only factors relating to husbandry and housing, which ultimately affect what the animals experience (resource inputs), but also their behavioural, physiological and pathological reactions to these experiences (animal-based outcomes). A wide variety of animal-based outcome measures (Tables A1, A2 and A3 in Appendix A) have been used to assess welfare, including unprovoked behaviours (e.g. stereotypy, aggression and overall behavioural repertoire), provoked responses (e.g. the reaction to a novel observer, object, and environment), and the physical appearance of individuals (e.g. starey coat, abnormal gait, chromodaccryea, and body score) (Leach and Main, 2008; Leach *et al.*, 2008). Government agencies and laboratory animal welfare groups have supported moves toward development of valid and feasible methods of assessing laboratory animal welfare. This has been backed up by various regulatory and advisory bodies in the UK, e.g. the Home Office review of the Local Ethical Review Process (Home Office, 2001) and recommendation 31 of the House of Lords Select Committee on Animals in Scientific Procedures (House of Lords, 2002). Post-operative and post-procedure monitoring schemes are now in widespread use as parts of the procedural assessment for animals undergoing licensed procedures, but such welfare assessments do not currently normally extend to assess the effect of housing and husbandry on laboratory animals despite this being considered to have profound effects on the quality of life for these animals (Figure A1 in Appendix A).

Basic information on laboratory animal assessment can be found on the website of Newcastle University (http://www.ahwla.org.uk/site/tutorials/HW/HW01-Title.html). This provides images of healthy animals and animals showing normal behaviours, and then contrasts these with images of animals with specific disease or pathology conditions and/or altered behaviours. The tutorial asks for 'answers' to questions about the animals' welfare status, and so trains the user in identifying problems and gauging their welfare significance. Additionally, the National Centre for the 3 Rs (Replacement, Reduction, Refinement) NC3Rs website has an introduction and further references on assessment methods (http://www.nc3rs.org.uk/category.asp?catID=22). For example, when evaluating an animal's physical condition, one should look at: body condition; changes in body shape; posture; fur and feathers; facial expression; skin; mucous membranes; eyes; ears; nose; mouth; and the tail.

Husbandry and housing can affect laboratory animal welfare in several ways - environmental conditions (e.g. temperature, humidity and sound levels), cage design (e.g. material, size, floor type and stocking density) (Baumans, 2005), the wide range of materials placed in animal cages (e.g. substrate, nesting material, shelters, gnawing material and other enrichments), routine husbandry practices (e.g. cage cleaning, handling and transport) and the establishment policies on monitoring animals and their environment (e.g. frequency of inspection, health records, standard operating procedures and health screening). The degree to which animal care staff are allocated and given time to care for laboratory animals is considered an important influence on animal welfare. In studies reported by Leach and Main (2008) and Leach *et al.* (2008) the number and type of animal care staff reported by the units coupled with the number of animals housed per unit enabled the calculation of the number of mice per full-time equivalent member of staff, which ranged from 24 mice per person to 3077 mice per person with the median being 570 mice per person. The inspection of the mice also varied greatly between units in both method and frequency, with 75% of units reporting that they inspected their mice daily. Of the remaining units, 11% reported inspections every 2-3 days, and 14% reported inspections only on a weekly basis. The reasons reported for not inspecting daily related to concerns that the disturbance caused by daily inspection would be detrimental to the health and welfare of the animals, for example causing infanticide if carried out too soon after animals were born. The vast majority (93%) of the units reported that they sent their animal care staff on training courses. Of the units that sent their staff on training courses, 57% assessed effectiveness of the training using feedback, continuous observation and formal assessment. All the units reported that they assessed the handling competence of their staff (see Tables A4 and A5 in Appendix A).

4.3 Laboratory animal application of animal based measures: humane endpoints

Practical strategies to assess and improve welfare are important for the use of animals in scientific research. However, the nature of their use presents some particular challenges and limitations. In most other applications, the main limiting factors in promoting good animal welfare are probably economic constraints. The market determines how high the price of animal products can be before the consumer chooses not to purchase, and this sets a limit to how many (costly) measures are implemented. Economic constraints also affect laboratory animal care, but in this area it is often the scientific use of animals that presents the most challenging hurdles. Not all research with animals results in animal suffering of course, but when the research requires that animals develop pathologies, or are exposed to pathogens or pharmaceuticals which cause disease or create cure, then this can potentially be accompanied by pain, injury, suffering or lasting harm. The more severe the disease or pathology caused, the more animal welfare will be compromised. Ethical concern over experiments causing animal pathologies has lead to the development of the concept of humane endpoints. These rely on 'non terminal' clinical signs to determine the point at which animals are euthanised (or sometimes treated) rather than awaiting death. (ILAR, 2000 http://dels-old.nas.edu/ilar_n/ilarjournal/41_2/index.shtml). An important component in the application of humane endpoints is the establishment of a welfare assessment protocol specific to the experiment in question. This should lay down how frequently the animals need to be

inspected, which parameters should be measured and criteria or a decision support tool for when to take action. In practice, these requirements usually translate into score sheets to be filled out by the person responsible for animal care and monitoring (Morton, 2000). Humane endpoints are applicable in a wide range of research and testing situations, including tumour research, infectious disease research, vaccine testing and tests of systemic toxicity. In many countries, a research project in which animals are expected to become severely ill will not be given ethical approval unless a protocol for humane endpoints is established. In regulatory testing of substances, the implementation of humane endpoints have lead to the replacement of the LD50 test (Lethal Dose 50, meaning the dose at which half of the animals die within 14 days of a single exposure) by tests in which most animals are euthanased at an earlier endpoint (OECD, 2002).

4.4 Zoo animals: animal welfare assessment in zoos and aquaria

Identifying effective and user-friendly welfare assessment tools is becoming a priority for zoos and aquaria. Over the past decade zoos have started to identify appropriate measures and processes to assess the welfare of captive zoo animals. In 2000, the Association of Zoos and Aquaria in the United States established an Animal Welfare Committee (AZA AWC) to address the increasing need for systematic and scientific assessment of zoo animal welfare. One of the first tasks of this committee was to initiate the compilation of AZA care guidelines, officially termed care manuals, for all of the zoo-held species. While only a small number of species have yet been covered by these Animal Care Manuals, at least 150 drafts are currently in various stages of the development and review process, and will be published over the next few years (see www.aza.org/animal-care-manuals/).

The goal of Animal Care Manuals is to assist in the development of consistent animal care practices across a wide variety of species regularly held in zoos and aquaria (Association of Zoos and Aquariums, 2009). Similarly, for farm animal assessment, Animal Care Manuals can only provide generalised resource-based and management-based welfare guidelines which allow zoos and aquaria to set the stage for good welfare (Barber, 2009) but naturally cannot provide an assessment of the actual welfare experience of individual animals.

A number of studies attempting to assess various aspects of individual zoo animal welfare and welfare outcomes based on variations in management or housing have been conducted (Wielebnowski *et al.*, 2002; Shepherdson *et al.*, 2004; Carlstead & Brown, 2005; Moreira *et al.*, 2007). These studies, while providing valuable information, have highlighted the complexity of animal focused assessments. To validate such measures properly, and to ensure their universal applicability, longitudinal data collection is usually necessary and large sample sizes are generally needed. This requires studies to be carried out across multiple institutions since common sample sizes for many species at any given zoo or aquarium are very small. In addition, the observation of zoo animals has to be conducted with only very limited access to the animals, and with respect to the fact that individual animals are often being viewed by zoo visitors (so keeper routines, animal visibility and visitor 'events' need to be taken into account when designing studies and experiments). These challenges combine to make it difficult to conduct animal outcome based studies on a regular basis to create welfare

assessments for the wide variety of species and individuals housed in zoos and aquaria. To provide appropriate assessments of individual well-being, animal-based welfare assessments should be used in combination with existing resource-based assessments. Animal-based measures can be integrated into ongoing welfare-monitoring by validating assessments based on keeper expertise and experience against quantitative measures. Scoring tools, once validated for a given species, may be employed at many institutions for long-term monitoring of individual animal well-being and can be integrated into day to day management practices.

4.5 Farm animal assessment - background to the 'Measures' used to assess animal welfare on farms

Animals differ in their experience, temperament and the way in which their genetic makeup interacts with their environment. The influence of management and the stockperson can also dramatically influence not only 'productivity' measures, but also the animals' experience of a particular situation. Thus, RBMs (like type of housing, allocation of resources, stocking density etc.) or management-based measures (like breeding strategies, health plans etc.) may be a poor guarantee of high levels of animal welfare in a particular situation.

Scientists have suggested that ABMs could provide valid indicators of animal welfare, since welfare is a characteristic of the individual animal, not just of the system in which animals are farmed. A comprehensive assessment of welfare should take as many different potential welfare 'factors' into account and should also be based on scientific knowledge of animal welfare where possible rather than on conjecture based purely on anthropomorphised positions. However, the 'societal position' on animal welfare should not be forgotten in the rush to promote robust scientific methods as, in the end, consumers make the purchasing decisions which influence what is purchased and so, what finally is the farming system (Kjaernes & Larvik, 2007)

Some single OBMs have been suggested as being capable of providing an integrated assessment of animal welfare including corticosteroids, acute phase proteins and longevity (Hurnik, 1990; Barnett & Hemsworth, 1990; Geers *et al.*, 2003). However, none of these single measures can cover all the dimensions of welfare. It seems probable that several measures are necessary to obtain a comprehensive view of any particular animal's welfare (Dawkins, 1990; Webster, 1997; Rutter, 1998). Recognising the difficulties of 'single measure approaches', grouped measures have been used to advise farmers (Sørensen, 2001): in 'branded' welfare certification schemes e.g. the Freedom Food Scheme (Main *et al.*, 2001); to compare systems to provide information during the creation of new legislation (Bracke *et al.*, 2002); and to check compliance with legislative requirements (Keeling & Svedberg, 1999).

Practically speaking, it may be possible to combine RBMs and ABMs. For example, if animals are lame (assessed using an ABM) it may be possible to predict lameness if the floor condition is poor (an RBM). However, the philosophy of ABM based assessment is that, if there is a measure from the animal which fulfils the conditions described above (practical, valid, repeatable, robust) then this should be used in preference to the RBM alone - because a given

floor condition may be very good for one animal but very poor for another. Additionally, it is clear that some simple questions, for example, 'is the animal thirsty' are not easy to answer in a farm situation. There is no currently feasible ABM for dehydration which can be carried out on the farm (though a blood measure could be carried out at the slaughterhouse, or it could be theoretically 'possible' to offer water to animals in a choice test).

In the last 30 years, there have been several moves in Europe, Australasia and North America to create assessment systems for farm animal welfare which use grouped assessment measures to address health and welfare criteria.

4.6 Recent developments in inclusion of animal assessment in farm assurance

Existing (privately owned, not Government-initiated) farm assurance standards are currently beginning to ask 'animal centred' questions, and it is apparent that repetitive resource based assessment can fail to fully answer questions about animal welfare. A number of current agricultural standards in use in Europe use basic 'welfare assessment tools' in that they incorporate the inspectors' view and interpretation of the severity of the issues observed. Some examples of the use of assessment standards can be found at:

- GLOBALGAP CCCP (Control points and compliance criteria) for beef cattle, sheep, dairy cows, pigs and poultry: http://www.globalgap.org

- Red Tractor Farm Assured Chicken Production: http://www.assuredchicken.org.uk/chicken/home.eb

Within these standards, most requirements are written as short descriptions of what is required and are almost all 'farm record' or 'resource' based (RBM). However, some of the existing standard 'clauses' are very close to being ABM tools - for example:

- *Are growing pigs kept in stable social groups?* (This could be answered by the stockman on the basis establishment of groups of pigs, but also could be answered by observing the social interactions between the animals.)

- *Where tail, flank, ear biting or fighting, which goes beyond normal behaviour becomes apparent, is an effective plan agreed…….?* (What is 'normal behaviour', what are the thresholds and what are the welfare implications for the animals of biting and fighting?)

- *Is the number of birds found dead on arrival at the slaughterhouse after transportation above 0.25%?* (Birds dying between catching at the farm and arrival at the slaughterhouse may reflect poor transport conditions or birds which were in a condition not suitable for transport in the first instance).

Across Europe, the use of assurance schemes in promoting higher levels of farm animal welfare is highly variable. 'Private' standards owners and assurance bodies can be flexible in

both driving standards upwards and in responding to local conditions. Private assurance schemes (not government driven), whether linked to NGOs (RSPCA Freedom Food, for example) specifically seeking to promote higher standards of farm animal welfare, or participants in retailer driven standards appear to offer current mechanisms for improving farm animal welfare. In some countries, scheme membership has now shifted from being entirely voluntary, to the current position where many producers now view membership to be an entry requirement (effectively non voluntary) to retailer shelves. In the overall assessment picture, animal welfare is perceived as a component of broader ranging assessments, which also contain environmental, COSHH, animal medicine use, work environment safety, food safety and retailer specific requirements. Moves towards animal welfare based assessment do have the potential to result in costs to producers. The European Commission estimated that the costs arising from the abolition of individual sow stalls was between €0.006 and €0.02 per kilogramme of pig meat (European Parliament, 2001) and an increase in egg production costs of a predicted 16% is anticipated by 2012 with the implementation of the ban on battery cages.

Assurance certification can only be applied to farms that operate within the applicable legal framework of the relevant country. As most assurance schemes make demands above (sometimes far above) the legal minimums, the net effect of assurance is to generally promote a voluntary raising of standards. Improvements in animal welfare can positively affect aspects of 'quality' through, for example, a reduction in bruising, bone breaks and blood spots (which translate to improved meat quality) or through increased disease resistance (Jones, 2001; Faure *et al.* 2003). If there is an increasing emphasis of assessment of ABMs, this may create a shift in the way that farms are inspected - moving the inspection from 'provision' to assessment based on assessing the 'quality of animal lives'. This is arguably a paradigmatic shift that sees the assessment of farm animals shift from a first position, 'protection against cruelty', to a 'focus on welfare' and then, from assessment of welfare to assessment of life quality (Buller & Morris, 2002; Buller & Roe, 2009). One of the possible effects of this shift could be a move towards the notion of farm animal welfare as a 'public good', something enriching the public sphere and so to be supported centrally rather than as a 'private initiative' articulated through privately owned commercial assessment systems.

4.7 The WelfareQuality® assessment system

Through collaboration between 41 institutes across Europe, the WelfareQuality® project (2004-2010) built on these principles to develop on-farm and slaughter assessment systems to address the areas of feeding, housing, health and disease and behaviour. WelfareQuality® created grouped assessment measures for pigs, poultry and cattle which addressed twelve health and welfare criteria which were (as with the Austrian ANI35) 'rooted' in a structure similar to that of the five freedoms (see Figure A2 in Appendix A) .

One of the aims of the Welfare Quality project was to develop tools for monitoring animal welfare from farm to slaughter i.e. including all the major stages of animal production - farm, transport, handling in the slaughterhouse and the slaughter process itself. The aims of these tools were that they could be used for a number of purposes:

- to allow inspection and scoring of farms, to inform consumers about the welfare status of the animals from which they buy products;
- to give 'advice' back to the farmer based on the data collected on the farm; and
- that the assessment protocols could be used by statutory bodies in their assessment of farms.

Key areas addressed by the WelfareQuality® assessment system were :

Hunger, thirst or malnutrition
This occurs when animals are denied a sufficient and appropriate diet or a sufficient and accessible water supply and can lead to dehydration, poor body condition and, for example, in chickens will be the final cause of death for sick or disabled birds which are not detected and then culled by the producer.

Physical comfort and security
Animals can suffer both short term and also chronic discomfort and can have problems lying down, getting up and standing. This can occur when they are kept in inappropriately designed housing or when they are transported in poorly designed or poorly ventilated vehicles.

Health- injuries
Animals can suffer physical injuries, such as mutilations, broken bones, bruises or skin lesions, due to: uneven or slippery flooring; enclosures with poorly maintained pens or gates; and fences with sharp metal, broken wood or concrete edges; and environments that promote aggressive behaviours between animals.

Health - disease
Animals can suffer diseases, for example, mastitis and metabolic disorders in cattle, lameness in pigs and poultry, and respiratory disease in calves. Disease is a 'common' experience for producers - and most take practical steps to reduce the incidence of disease through biosecurity measures, vaccination and housing environment control - but when disease does occur, it is the step which the producer then makes which can profoundly affect the welfare of animals - is disease rapidly diagnosed and then effectively managed? Poor hygiene, irregular monitoring and insufficient treatment speeds can amplify the effects of disease on animals.

Pain (not related to injuries or disease)
In addition to suffering pain from injuries and disease, animals can experience intense or prolonged pain due to inappropriate management, handling, slaughter, or invasive management procedures (e.g. castration, dehorning, tail docking) and as a result of aggressive encounters between animals, and also due to the use of goads (electrical or not) used to move them.

Normal/natural social and other behaviours

Animals can be denied the opportunity to express natural, non-harmful, social behaviours, such as grooming, social licking and huddling for warmth. Separating females from their offspring very early in production, and preventing sexual behaviour are examples of more established restrictions on 'normal behaviour' which are common in farming systems. Animals can also be denied the possibility of expressing other intuitively desirable natural behaviours, such as exploration and play. The denial of these possibilities may lead to abnormal and/or harmful behaviours such as tongue rolling in cattle, bar biting and tail biting in pigs and feather pecking in chickens.

Human-animal relationships

Poor relationships with humans can be reflected in increased avoidance distances and fearful or aggressive animal behaviours. This can occur due to inappropriate handling techniques (e.g. slapping, kicking and the use of electric goads), or when producers, animal transporters or slaughterhouse staff are either insufficiently skilled or possess unfavourable attitudes towards animals.

Positive emotions and avoidance of negative emotions

Animals can experience emotions such as fear, distress, frustration or apathy when they are kept in inappropriate physical or social environments (e.g. where there is a lot of mixing of groups of animals, or not enough space to avoid an aggressive partner). These emotions can be reflected in behaviours such as panic, flight, social withdrawal and aggression and behavioural disorders. Poor management routines and a lack of environmental stimulation may prevent animals from expressing positive emotions. These are not easy to assess but may be reflected in certain behaviours, such as play (especially in young animals) and positive social interactions (social licking). Some (including many animal scientists) consider that promoting 'positive emotional states' in animals is beyond the normal requirements for farmers who are business people trying to make a living from rearing animals, whereas others consider these exact reasons to be good justification for producers to show 'care' for all aspects of the needs of the animals which provide their livelihood. Practically speaking, 'care' for the positive needs of animals could be considered an indicator of the ability of the farming systems to consider the whole picture of what farmed animals need to live and to thrive.

Table 1 gives an example of the Welfare Quality® system providing a description of the assessment methods used for beef cattle.

Table 1. An example of the range of measures for cattle for grouped measures proposed in the Welfare Quality® farm assessment system.

	Welfare criteria	Measures
Good feeding	1. Absence of prolonged hunger	Body condition score
	2. Absence of prolonged thirst	Water provision, cleanliness of water points, number of animals using the water points
Good housing	3. Comfort around resting	Time needed to lie down, cleanliness of the animals
	4. Thermal comfort	As yet, no measure is developed
	5. Ease of movement	Pen features according to live weight, access to outdoor loafing area or pasture
Good health	6. Absence of injuries	Lameness, integument alterations
	7. Absence of disease	Coughing, nasal discharge, ocular discharge, hampered respiration, diarrhoea, bloated rumen, mortality
	8. Absence of pain induced by management procedures	Disbudding/dehorning, tail docking, castration
Appropriate behaviour	9. Expression of social behaviours	Agonistic behaviours, cohesive behaviours
	10. Expression of other behaviours	Access to pasture
	11. Good human-animal relationship	Avoidance distance
	12. Positive emotional state	Qualitative behaviour assessment

Training assessors to carry out the measures

- Training in scoring is by means of on-farm visits and examination of photographs, video and by assessment of animals during farm visits.
- The inspectors will require to be assessed during the training course until they develop a uniform scoring.
- Assessors will also be asked to periodically carry out a validation or reference audit to check that they continue to score in a repeatable way.

4.8 Turning assessments into 'scores'

Once measures have been carried out on a farm, it is possible to create a range of 'scores'. The first level of scores would be the results for each measure, provided initially as 'benchmarking' or 'initial position' scores, which tell the producer how he is positioned at the very start of the assessments, and also allows him to compare himself with some known baseline values and, also, with his farming peers. The individual scores for each measure made can be fed back to the farmer, so that he can see how he did, and also to support any areas that he might wish to develop or focus on (if he has a problem). For example, he might find that the assessment shows that he has an unusually high number of tail bitten pigs in a particular group (when compared to his previous assessments). He might use the periodic information provided by an assessment as a barometer for any changes he chooses to make - if they cause reductions in the 'area of concern', then he can use this information to assess

the degree of 'success' that he has achieved - and conversely, if there is no improvement or even an increase, he may be able to use the risk factor assessment to target other areas of change. Von Borell *et al.* (2001) created hazard analysis and critical control points (HACCP) based recording systems for pig housing, with the potential to feed back results of the effects of altered management at these critical control points. The concept of feeding back individual 'non compliant' assessment points is the basis of all existing farm assurance and certification schemes, and so this approach is familiar to many farmers, and is seen as the basis of a reasonable approach to ensuring compliance with a recognised 'standard'.

The second level of 'scoring' would be to combine the individual assessment results - perhaps into linked groupings - for example, in the area of housing. A combined score for related areas is possible but only if:

- The individual measures can be combined to give aggregate scores which are considered meaningful, and credible, by both the producer and the consumer.

- The process of combining scores must not devalue the overall meaning of the assessment information, for example, by compensating a very poor score in one area, with a number of 'ok' scores in other areas. If this occurs, then the power of discrimination can be lost, and the credibility of the combined score brought into doubt. Spoolder *et al.* (2003) indicated that when welfare scores are simply summated to give an overall score, a welfare disadvantage can be compensated for by a number of minor advantages, but effect can be limited (or even eliminated) if minimal requirements are set below which specific score cannot be 'permitted' (Botreau, 2007b and c).

Once raw 'values' for measures have been made on the farm, the 'natural' next step is to ascribe different weighted values to these measures, to give impact factors for each measure with respect to animal welfare (Bracke *et al.* 2002). This type of weighting system is seen in the Austrian TierGerechtheitsIndex (TGI) (Bartussek, 1999b). Weighted sums of scores appear intuitive and the principle is usually readily understood by users.

4.9 Conclusions

Research scientists have suggested that ABMs could provide valid indicators of animal welfare, since welfare is a characteristic of the individual animal, not just of the system in which animals are farmed. The sorts of questions which are being asked are:

- are the animals properly fed and supplied with water?
- are the animals properly housed?
- are the animals healthy? and
- can the animals express a range of behaviours and emotional states?

In current work being carried out around the world, it is clear that a 'bringing together' of expertise and the start of a consensus amongst scientists is taking place, and that it is considered possible to use ABMs to provide information about the quality of life of farmed animals and of how well animals are being farmed within the existing 'systems' which are prescribed by legislation and established farming practice. The tools currently being developed in Europe in ABMs or OBMs following on from the Welfare Quality project evolve the trend toward inclusion of assessment techniques which reflect what can be measured 'on the animal'. Whilst accepting that many areas of animal assessment are complex, and some are very subtle, it seems likely that some ABMs are likely to find their way into farm assessment schemes and, to a certain degree, into the farm animal legislation. This raises some fears in the 'community', if ABMs are of use, how far will people want to go with them? Should (or even could) they replace some of the prescriptive RBMs found in existing legislation? Some people appear to be concerned that the good work already carried out in animal protection which is now enshrined in existing legislation could be 'swept away' on a wave of enthusiasm for ABMs. The inclusion of ABMs is a ripple of evolution (not a wave of revolution) and the stability and conservatism of the legislative and assessment community will ensure that, if ABMs do begin to make their way into farm assessment methods and the legislation, then this will take place gradually. In studies carried out within the Welfare Quality® project (Keeling & Bock, 2007), it was found that participation in animal schemes affects farmers' definition of animal welfare, the importance they attach to it, and their willingness to further increase animal welfare standards. Three groups of farmers were identified: farmers who participate in specific animal welfare schemes or organic production, farmers who participate in quality assurance schemes, and those who are part of no scheme at all. The first group of farmers was found to be the most concerned about animal welfare and were driven by ethical consideration and their ambition to carry out alternative agricultural production with the potential to receive a premium price as an important potential incentive. On the one hand, farmers hope that animal-friendly products will offer opportunities for new markets, but they may also feel threatened by the potential import of cheaper food and the resulting competition on price. For all farmers, a secure market is an important driving factor. Some farmers do not participate in specific animal welfare schemes because they have no faith in the financial benefits promised and see only an increase in bureaucracy and fear a loss of their independence to manage.

4.10 What further work is likely?

A number of farmed animal species are not represented fully in existing work on ABMs. For example, fish, small ruminants, poultry other than laying hens and broilers and animals farmed for their fur. These species have received a large amount of focussed scientific study but, to date, there has not been the collaborative work to combine knowledge in these species to produce animal welfare assessment systems. At present, the use of ABMs in the commercial setting and on a significant scale outside of controlled studies and for a realistic period has not taken place. ABMs are, in general, 'complementary to existing frameworks until proven otherwise' and so, if the use of ABMs in commercial assessment systems can be initiated, most likely in collaboration with existing certification and assurance bodies, then it will be possible to determine whether the claims that ABMs offer real value to the consumer

and producer are realistic. The consumer could potentially benefit, through choice backed up by product information (and potentially labelling), and the producer may find that he/she can make better, more 'informed', management decisions.

The protocols for animal welfare assessment from the Welfare Quality® project have been tested and are being applied by a number of researchers within Europe. A drawback to the current protocols is that, although they are comprehensive, they are relatively time consuming to implement. For this reason, within the UK, for example, there is a project involving Bristol University, RSPCA Freedom Food and the Soil Association to test the application of (modified) protocols to farm assurance schemes by undertaking farm-based assessments and feeding results back to farmers and others in the supply chain.

4.11 Summary

Existing animal welfare assessment schemes on farm, in laboratories or in zoos, tend to assess welfare by examination of the provision of housing or resources (RBMs), rather than looking at outcomes for the animals themselves (OBMs). Scientists have suggested that OBMs could provide valid indicators of animal welfare, as welfare is a characteristic of the individual animal, not just of the system in which they are kept. The sorts of questions being asked are: *Are the animals properly fed and supplied with water? Are the animals properly housed? Are the animals healthy? Can the animals express a range of behaviours and emotional states?* It is now increasingly being felt that these questions are best addressed by a combination of both animal based (OBM) measures and resource based (RBM) measures and many people can see that a balanced mixture of animal and resource based measures are likely to be the way in which farm assurance may move forward in the future. Many farms, zoos, aquaria and other facilities are keen to identify welfare monitoring tools that can be readily applied and allow for rapid assessment to provide timely feedback for management. To implement effective use of ABMs, it is necessary to adopt a number of logical steps. First, it is necessary to measure the 'level' of welfare conditions (using both RBMs and OBMs), then analyse the risk factors - the variables which may influence the development of welfare 'issues', then inform the producer, animal keeper or zoo (and perhaps the user or purchaser) what has been found. Finally, they should support management decisions necessary to create improvements in welfare. When OBMs have been made, it may then be possible to create a range of 'scores' which can be combined to give aggregate scores, which can be presented to the producer (animal carer, zoo etc.) and even to the consumer. The Welfare Quality tool has already the capacity to give an aggregated score to a farm through use of a webtool based calculation tool which returns a score if farm data is input. This requires the attribution of weighted values to the measures, to assess the impact of each measure with respect to animal welfare. A multi-pronged strategy involving various RBMs and OBMs is most likely to provide the capacity for overall welfare monitoring of animal welfare in the future. Such a strategy may include regularly updated species care guidelines, accreditation standards, longitudinal studies and the development of practical assessment tools (including potentially automated systems using cameras) for welfare monitoring.

5. Valuing farm animal welfare – a review of methods and applications

5.1 Theoretical background

There are two kinds of benefits associated with farm animal welfare. First, private animal welfare benefits (PAWB) are the utility that consumers of high animal welfare status products derive from the perceived better taste, food safety, healthiness or nutritional value (Rolfe, 1999; Miele and Evans, 2005; European Commission, 2007) of these products plus any animal productivity gains from better welfare. Second, social animal welfare benefits (SAWB) are the benefits that are experienced by those who care about the livestock production process itself due to environmental, ethical, or religious reasons, from knowing that animals are kept to high welfare standards plus any additional social benefits associated with, for example, environmental improvements – such as lower green house gas emissions. For many people, knowing that others cannot consume products from animals that experienced poor welfare, and thereby cannot support the existing welfare standards that they disagree with, creates utility (Bennett, 1995).

SAWB are non-excludable and non-rival as livestock producers who keep their animals at superior welfare levels cannot arrange to be paid for their services from everyone who benefits from knowing about this. The optimum supply of animal welfare is, therefore, financially unprofitable in the private market, even if it is overall economically viable to society. Thus, SAWB can be regarded as a public good, reflecting the wider social value placed upon animal friendly production (Henson and Beard, 1994), and their existence leads to calls for Government intervention such as legislation to improve farm animal welfare both on, and off, the farm.

5.2 Methods for valuing farm animal welfare benefits

This section reviews and appraises different approaches that can be, and have been, used to value the aforementioned kinds of farm animal welfare benefits. They are categorised into 'direct', 'indirect', and 'auction' methods. Their merits, limitations and applicability to the valuation of farm animal welfare are listed in Table 3 below.

5.2.1 Direct methods

These kinds of methods are often used when market data are unavailable or unreliable because they allow researchers to create and present hypothetical scenarios to respondents and ask them to state their preferences directly in terms of hypothetical markets or payments.

The contingent valuation method

The contingent valuation method operates by setting a hypothetical market for the provision or increase of a non-market good. Survey questions are used to find out how much people are willing to pay for specified increases in the non-market good, or willing to accept for specified

reductions of it (Mitchell and Carson, 1989). This way, the total economic value of a scenario or policy can be established. A number of different value elicitation formats can be used for individuals to express their maximum willingness to pay (WTP) such as: the open-ended question format, which is particularly useful for exploratory pilot studies (Bennett and Tranter, 1998) but does not provide an incentive to reveal the true maximum WTP; the single-bounded dichotomous choice format (Hanemann, 1984) which is a straightforward and simple to answer elicitation format (Loomis and Helfand, 2003) that is incentive compatible; the double-bounded dichotomous choice format (Carson, 1985; Hanemann, 1985) which increases the efficiency of the resulting welfare estimates compared to the single-bounded format (Alberini *et al.*, 2003) but is not incentive compatible (Leon and Vazquez-Polo, 1998); or the payment card (Mitchell and Carson, 1989) which provides more efficient statistical information (Champ *et al.*, 2003) by showing respondents *k* bid amounts and asking them to choose their maximum WTP amount.

The value estimates in contingent valuation studies can be affected by a number of biases, such as: scope insensitivity if respondents fail to offer a significantly higher WTP for larger amounts of a good (Burgess *et al.*, 2003); information bias if the value estimates depend on the information that was provided to respondents; anchoring if the valuation depends on the first bid presented (Pearce and Barbier, 2000); or hypothetical bias if the values elicited in the context of the contingent valuation study differ from those elicited in a real world context. The existence of the latter in contingent valuation studies has been well documented in the literature (e.g. Harrison and Rustrom, 2008).

Numerous contingent valuation studies have been carried out to value farm animal welfare benefits over the last 15 years (e.g. Bennett and Larson, 1996; Bennett, 1997; Rolfe, 1999; Bennett and Blaney, 2002; Bennett *et al.*, 2002; Lusk and Fox, 2002; Bennett and Blaney, 2003; Burgess *et al.*, 2003; Burgess and Hutchinson, 2005; Glass *et al.*, 2005; Moran and McVittie, 2008; Taylor and Signal, 2009; Kehlbacher, 2010; Nocella *et al.*, 2010). The contingent valuation method is appropriate to value both PAWB and SAWB. Due to its hypothetical nature it further can be used for *ex-ante* analysis of the possible introduction of legislation or codes of practice. However, the fact that the WTP questions are hypothetical and that animal welfare is an emotive subject may cause respondents to express their attitudes rather than their preferences and, in this way, overstate their WTP causing value estimates to become inflated.

The choice experiment method

The choice experiment method is an attribute-based survey method for measuring benefits. It has its roots in Lancaster's characteristics theory of value (Lancaster, 1966), which postulates that preferences for goods are a function of the characteristics or attributes possessed by the good rather than the good *per se*. Accordingly, a good can be described by those of its attributes that generate utility or disutility.

In a choice experiment, respondents are presented with numerous sets of alternative combinations of attributes and asked to choose their most preferred alternative thus

revealing the trade-offs they are willing to make between the different attributes. One of the attributes is typically a monetary cost or benefit attribute that serves as a proxy for the marginal utility of income and allows the calculation of respondents' WTP for the remaining attributes (Louviere *et al.*, 2000). Accordingly, the main advantage of the choice experiment method is that, apart from producing the total economic value of an alternative, it allows investigators the discovery of how WTP is built up from different components of perceived value.

Unlike the contingent valuation method, choice experiments suffer less from respondents behaving strategically because there is no clearly identifiable agenda in a sequence of choices as almost all levels of attributes change from one choice set to another (Veisten and Navrud, 2006). Choice experiments further appear to be less prone to hypothetical bias because they closely mirror actual consumer purchasing situations (Lusk and Schroeder, 2004). In addition, hypothetical bias in marginal WTP is not a major concern in choice experiments if the main interest of a study lies in the relative magnitude of the estimated marginal WTPs. Arguably, potential biases may not differ systematically between attribute levels making a comparison between two attribute levels less vulnerable to hypothetical bias (Carlsson *et al.*, 2007a). However, cognitive difficulties tend to be more accentuated in choice experiments due to their rather repetitive nature and because respondents are required to evaluate large, lengthy and complex choice sets (DeShazo and Fermo, 2002). This means that choice consistency can be affected by learning (List *et al.*, 2006) or subject fatigue (Johnson and Desvousges, 1997). Another problem of the choice experiment method is that some respondents display lexicographic preferences by choosing alternatives based on the one attribute that is most important to them. This violates the underlying assumptions of strict convexity and continuity (Lockwood, 1996). Strict convexity implies that indifference curves have no flat spots so that there will be only one optimal choice on each budget line. Continuity implies that for any two attributes it will always be possible to accept an amount of the second attribute in order to compensate for the loss of a small amount of the first attribute.

A variety of choice experiment studies have been carried out in the last 10 years to value different aspects of farm animal welfare (Lusk *et al.*, 2003; Carlsson *et al.*, 2005a and b; Lagerkvist *et al.*, 2006; McVittie *et al.*, 2006; Carlsson *et al.*, 2007a and b; Lusk *et al.*, 2007; Liljenstolpe, 2008; Tonsor *et al.*, 2009; Kehlbacher, 2010; Morkbak *et al.*, 2010). Like the contingent valuation method, choice experiments are suitable for valuing both PAWB and SAWB and, due to their hypothetical nature, they can be used for ex-ante analysis of possible legislation or codes of practice. In conjunction with a welfare score, the choice experiment method can be used to obtain marginal values of farm animal welfare benefits and to investigate respondents' trade-offs between the welfare of different animal species (Kehlbacher, 2010).

The non-hypothetical choice experiment method

The non-hypothetical or real choice experiment method is an extension of the choice experiment method. After having completed all choice sets, one of the chosen scenarios is randomly selected. This selected scenario is the binding scenario and participants purchase

the product they chose in this scenario. The non-hypothetical choice experiment method is incentive compatible (Alfnes *et al.*, 2005) but in the context of valuing farm animal welfare it is only suitable for valuing PAWB. A study using this method was carried out to value the welfare of farmed salmon (Olesen *et al.*, 2010).

5.2.2 Indirect methods

Indirect methods examine available markets for private goods which are related to the non-market good. The value of the latter is reflected indirectly in consumer expenditure or in the prices of the marketed goods.

Hedonic pricing

The hedonic pricing approach (Rosen, 1974), like the choice experiment method, is consistent with Lancaster's theory of consumer demand (Lancaster, 1966). A good is regarded as a set of characteristics and its value is a function of those characteristics. Holding all characteristics constant except for one, and comparing this with the differences in prices, it becomes possible to estimate the marginal value of the characteristic in question. The main strength of hedonic pricing is that value estimates are based on real economic commitments. However, basic problems include omitted variable bias, multicollinearity, choice of functional form, market segmentation and restrictive market assumptions (Kjær, 2005).

The underlying assumptions of the hedonic pricing method are that consumers have perfect knowledge in relation to the quantity of the non-market good; all buyers are able to go to the utility maximum position; and the market is in equilibrium. Further assumptions include weak separability meaning that the marginal rate of substitution between two goods in an individual's utility function is independent of the quantities of all other goods, and weak complementarity implying that when the cost of consuming the non-market good is zero, the marginal WTP of the good is zero. It is because of the two latter assumptions that hedonic pricing cannot value non-use values (Hitzhusen, 2007).

A study using this method to value farm animal welfare was carried out by Baltzer (2004) who provided an estimate of the contribution that the characteristic 'free-range' makes to the price of eggs. The hedonic pricing method is only to a limited extent useful for valuing farm animal welfare. First, due to the assumptions of weak separability and weak complementarity, the hedonic pricing method cannot capture SAWB. Second, because currently there are only a few livestock products on the market that have attributes pertaining to animal welfare, in the eyes of consumers the market for farm animal welfare products is not in equilibrium thus violating the underlying assumption of the hedonic pricing method. Finally, any analysis of existing animal welfare friendly production practices is necessarily ex-post and may be of little help with ex-ante analysis of possible new legislation or codes of practice, except only by analogy (Bennett, 1995).

Travel cost method

The travel cost method is generally used to value environmental goods and services in which the beneficiary has to travel to enjoy or experience them. Different individuals are willing to spend different amounts of time and travel to get to different sites. These implicit prices reflect the value in utility they get from making the trip (Hanley *et al*., 2001). Like hedonic pricing, the travel cost method relies on the assumption of weak complementarity, which implies that when the cost of consuming the services of the non-market good is zero, then the marginal WTP of the good is zero, and it therefore can be used only to value use values. Since the method relies on there being a direct relationship between travel cost and the environmental effect, the method's applicability to the valuation of farm animal welfare benefits is, therefore, limited. Accordingly, no study using this method to value farm animal welfare was found in the literature. A conceivable study scenario would involve people travelling to visit a farm shop selling high welfare status products. It is difficult, though, for the researcher to establish the true purpose of such a journey and disentangle the farm animal welfare benefits from the benefits related to the enjoyment of the trip and from the benefits of buying products not derived from animals.

Averting expenditure approach

The averting expenditure method uses changes in spending on goods that are substitutes for the non-market good in question. The motivation is that individuals' utility is negatively affected by, for example, exposure to pollution. Consequently, individuals undertake averting action by purchasing goods and services to reduce their exposure. Once the general environment is improved by some policy initiative, the individuals can spend less on these substitute goods. It is this decline in averting expenditure that gives an indirect estimate of the individuals' WTP for the associated incremental benefits provided by the decrease in pollution. A problem of this valuation approach is that if there is jointness in the production of the output, that is averting expenditure yields benefits beyond those associated with the improvement in air quality, this method will overestimate the benefits of the environmental effect (Callan and Thomas, 2004).

In the literature, no study using the averting expenditure approach to value farm animal welfare benefits was found. The method's applicability to the valuation of farm animal welfare is limited because it requires people to be informed about the level of welfare in livestock production. For example, if a comprehensive and credible system of welfare labelling existed, consumers would be able to avoid consuming products with lower animal welfare status. Once new welfare legislation is introduced, the reduction in consumer spending on high welfare status products would reflect consumers PAWB. Being an indirect method that observes market behaviour, the method is not appropriate to value SAWB.

The subjective well-being method

Subjective well-being is a broad category of phenomena encompassing people's emotional responses, domain satisfactions, and global judgements of life satisfaction (Diener *et al*., 1999).

The components of subjective well-being are depicted in Table 2. The method assumes that subjective judgements are a trustworthy representation of how well life is going for that person (Dolan and Peasgood, 2006), and that individuals understand and respond to subjective questions in similar ways. The latter assumption allows for interpersonal comparisons of individual welfare which are a necessary requirement for the construction of a social welfare function (Ferrer-i Carbonell, 2002).

Table 2. Components of subjective well-being

Affective component		Cognitive component	
Positive affect	Negative affect	Life satisfaction	Domain satisfaction
Joy	Guilt and shame	Desire to change life	Work
Elation	Sadness	Satisfaction with current life	Family
Contentment	Anxiety and worry	Satisfaction with the past	Leisure
Pride	Anger	Satisfaction with the future	Health
Affection	Stress	Significant others' view of one's life	Finances
Happiness	Depression		Self
Ecstasy	Envy		One's group

Source: adapted from Diener *et al*., 1999.

Subjective well-being can be measured in a variety of ways including: global self-reports in surveys that ask respondents to evaluate their lives as a whole or some aspect of it (van Hoorn, 2007); experience sampling where personal digital assistants are used to ask people at random times during the day to rate different feelings (Dolan and Peasgood, 2006); day reconstruction methods where respondents are asked to divide the previous day into a number of episodes and then to rate different elements of affect during those activities on a scale (Dolan and Peasgood, 2006); the U-index (Kahneman and Krueger, 2006) which measures the proportion of time an individual spends in an unpleasant state; and the life satisfaction approach which asks respondents about their life satisfaction in general or with respect to various domains (van Hoorn, 2007). By not asking respondents to value the non-market good directly, but to evaluate their general subjective well-being, the method avoids some of the difficulties associated with stated preference methods. It is cognitively less demanding (Frey *et al*., 2004), and it is subject to fewer measurement biases such as elicitation bias, embedding (Dolan and Metcalfe, 2008), focusing effect, hypothetical bias, and strategic bias. It further avoids problems with respondents not considering the effect of their budget constraints and substitutes, and it does not rely on respondents' ability to consider all relevant consequences of a change in the provision of a public good (Stutzer and Frey, 2010).

The main drawback of the method is that it does not produce monetary value estimates that can be included in a cost-benefit analysis. In addition, while the subjective well-being method does fairly well on the various dimensions of validity, its reliability has been found to be modest (Krueger and Schkade, 2007; van Hoorn, 2007).

To date, no study has been carried out to value farm animal welfare benefits using the subjective well-being method although the method would be suitable to value SAWB as well as PAWB. However, no monetary value estimates that can be used in farm animal welfare policy decision making have been provided, and the analysis is limited to existing animal welfare friendly production practices and thus offers little help with *ex-ante* analysis of possible legislation or codes of practice.

5.2.3 Auction-based methods

Auction-based methods such as the second price Vickrey auction create an artificial choice situation but choices made by participants are real (Buzby *et al.,* 1998). Participants are endowed with a good that is typically an existing product in a given product category. A different good in the same product category is put up for sale and participants in the market reveal the value that they place on the good for sale by indicating the highest amount of money they are willing to pay to exchange their endowed good for the good on sale. The highest bidder has to give up the endowed product and pay a price equal to the second highest bid. In this way, a person cannot be made better off by misrepresenting his or her actual value. The difference in a respondent's bids for products that differ in only one attribute makes it possible to derive the WTP for this particular attribute (Jaeger *et al*., 2004).

A key advantage of the experimental auction method is that it is an incentive compatible method of eliciting WTP because participants' weakly dominant strategy is to bid their own reservation price (Alfnes and Rickertsen, 2003). In addition, the introduction of a budget constraint can help overcome inconsistencies such as scope insensitivity as often found in stated preference studies (Buzby *et al*., 1998). However, experimental auctions are usually conducted in a laboratory setting, which means that samples of consumers are relatively small in size and tend to be locally recruited (Alfnes and Rickertsen, 2007). The auction method is an unusual market mechanism as, in reality, the decisions at the point of purchase are often made with far less involvement, and consumers are not bidding for a limited stock (Enneking, 2004). The presentation format and auction protocol can affect prices in ways that make it difficult to compare results and limits their broad applicability (Melton *et al*., 1996).

Auction-based methods can be used to value farm animal welfare benefits by obtaining WTP for food items with attributes pertaining to animal welfare. The food items that have been used in past studies include roast beef and ham sandwiches (Dickinson *et al*., 2003), yoghurt (Napolitano *et al*., 2008), eggs and pork chops (Norwood and Lusk, 2008), cured ham (Gracia *et al*., 2009) and raw organic beef (Napolitano, 2009). Because farm animal welfare can only be valued in the form of a product attribute, auction-based methods are suitable only to value PAWB while SAWB cannot be captured using them.

Table 3. Valuation Methods

Method	Merits	Limitations	Applicability to farm animal welfare valuation
Contingent valuation	can measure both use and non-use values; obtains total economic value	hypothetical bias; scope insensitivity; information bias; anchoring	can be used to value both PAWB and SAWB; can be used for ex-ante analysis of policy measures
Choice experiment	can measure both use and non-use values; obtains marginal value estimates of different attributes	hypothetical bias; cognitively demanding	can be used to value and distinguish between PAWB and SAWB; can be used for ex-ante analysis of policy measures; can value species trade-offs
Non-hypothetical choice experiment	obtains marginal value estimates of different attributes; incentive compatible	cognitively demanding ; cannot measure non-use values	can value only PAWB because animal welfare has to be a product attribute
Hedonic pricing	observes real behaviour	cannot measure non-use values; can measure only characteristics of which individuals are aware; market has to be in equilibrium	animal welfare product market is not in equilibrium can value only existing welfare standards; cannot value SAWB
Travel cost method	observes real behaviour	elicits preferences only of those who travel; cannot measure non-use values; difficult to establish true purpose of the journey	can value only existing welfare standards; difficult to disentangle animal welfare benefits from other benefits of the trip; cannot value SAWB
Averting expenditure	observes real behaviour	cannot measure non-use values; can measure only non-market goods of which individuals are aware; requires policies to have caused changes in the provision of the non-market good	can value only existing welfare standards; cannot value SAWB
Subjective well-being	observes real behaviour; no focus effect or strategic bias; questions are easy to answer	cannot measure non-use values; can only value goods that affect subjective well-being; does not produce WTP estimates	can value both SAWB and PAWB but only of existing welfare standards; people need to be aware of, and affected by, farm animal welfare standards
Auction-based methods	involves real choices; incentive compatible; participants observe budget constraint	cannot measure non-use values; small and local sample; laboratory setting	can value only PAWB because animal welfare has to be a product attribute

5.3 Review of the existing literature

This section critically reviews past farm animal welfare valuation studies with regard to their theoretical and construct validity. Table 4 is an extension of a listing of farm animal welfare valuation studies initially provided by Lagerkvist and Hess (2011). It reports the studies covered by this review including information about the method used, country of study, sample size, the type of animal whose welfare is being valued, the good being valued and WTP estimates in both £ and, where it is possible to derive them, as a %.

Studies valuing farm animal welfare benefits have used several different ways of eliciting respondents' preferences for farm animal welfare. One way has been to elicit respondents' WTP for a food product with one or several attributes pertaining to animal welfare using the choice experiment method (e.g. Lusk *et al.*, 2003, 2007; Liljenstolpe, *2008*; Tonsor *et al*., 2009; Morkbak *et al*., 2010; Anonymous, 2010) or the contingent valuation method (e.g. Bennett, 1997; Nocella *et al*., 2010; Taylor and Signal, 2009). In addition, due to the nature of the methods, animal welfare is always valued in this way if experimental auctions (e.g. Dickinson *et al*., 2003; Napolitano *et al*., 2008; Gracia *et al.*, 2009; Napolitano, 2009) or non-hypothetical choice experiments (e.g. Olesen *et al*., 2010), are used. The advantage of valuing farm animal welfare as a product, or production process attribute, is that respondents are accustomed to encountering and thinking of animal welfare in this way. This, in turn, may increase the reliability of their answers to the valuation questions. However, respondents who for ethical, religious, health, or other reasons do not consume the products in question cannot credibly reveal their preferences in this way. This is an important shortcoming if the objective of the study is to make policy recommendations. Indeed, a fundamental premise of welfare economics is that public policy decisions should reflect the preferences of those who will be affected by them.

Another way of uncovering respondents' preferences for farm animal welfare has been to elicit respondents' WTP for farm animal welfare legislation or production regulation. Burgess *et al.* (2003), Burgess and Hutchinson (2005) and Glass *et al.* (2005) valued the introduction of farm animal welfare regulations using a food tax as a payment vehicle. Arguably, this is not a realistic payment vehicle because of the regressive effect a food tax would have on income. Other studies eliciting respondents' WTP for farm animal welfare legislation have used payment vehicles that are incongruous with the benefits accruing from farm animal welfare legislation. Price increases as used by Bennett (1996, 1997) and Bennett and Blaney (2003) or increases in food expenditure as used by Bennett and Blaney (2002) and Bennett *et al*. (2002) are inappropriate payment vehicles because such legislation benefits not only those who purchase livestock products and so these valuations fail to capture the full economic value associated with improvements to animal welfare.

An appropriate payment vehicle that is credible and conforms with the type of benefits produced by legislation is an increase in or introduction of a general tax, given that respondents pay tax. This approach has been used in a number of valuation studies (Bennett

and Larson, 1996; Moran and McVittie, 2008; Kehlbacher, 2010). It allows the measuring of both consumers' PAWB and also the SAWB of the total population, including non-consumers.

Few studies have attempted to make a distinction between PAWB and SAWB when valuing farm animal welfare benefits. This distinction is useful because policy makers can address PAWB by introducing a labelling scheme while SAWB, having public good characteristics, can be addressed through legislation. Accordingly, Carlsson *et al.* (2007b), Tonsor *et al.* (2009) and Kehlbacher (2010) compared the benefits achieved from welfare labelling and welfare legislation so as to isolate SAWB from the PAWB. While Carlsson *et al.* (2007b) and Tonsor *et al.* (2009) accounted only for the preferences of consumers, Kehlbacher (2010) used a different methodology in order to include the SAWB of non-consumers in the analysis as well. Concordantly, the three studies found PAWB to exceed SAWB. The authors see the reason for this in a notion by Hamilton *et al.* (2003) who argued that preferences for public goods also may include an aversion against a loss of options. Accordingly, the authors suggested that, if consumers are provided with adequate labelling that allows them to buy high welfare status products according to their preferences, there is no economic support justifying legislation because the private loss of option values is offsetting any public good benefits of the legislation.

As regards the extent of farm animal welfare improvements that were valued, only one study assessed the total value of farm animal welfare by asking respondents for their maximum WTP to address all their farm animal welfare concerns together (Bennett and Blaney, 2003). Most studies valued welfare improvements for one or occasionally several animal species (e.g. Carlsson *et al.*, 2007a). In doing so, they elicited respondents' preferences for the introduction of, or changes in, one or several husbandry practices. For example, Bennett (1997) valued the ban of battery cages in egg production, Tonsor *et al.* (2005) valued the ban of gestation crates in pig production, McVittie *et al.* (2005) asked respondents to consider different levels of stocking density, ventilation, period of darkness, and percentage of flocks failing food pad lesion standards to improve broiler welfare, Lagerkvist *et al.* (2006) elicited preferences for attributes relating to the housing system, castration techniques, tail docking and restraint to value pig welfare and Carlsson *et al.* (2005a) asked respondents to consider attributes relating to fodder, outdoor production, transport and varying the choice of breed to improve the welfare of broilers, beef cattle and pigs.

Table 4. A summary of the findings of animal welfare valuation studies, 1996-2010

Study	Method	Country	Sample size	Animal type	Valued good	Mean WTP (£)*		WTP (%)
Anonymous (2010)	Choice experiment	CA	623	pigs	pork chops	30.02	per kg	-
Baltzer (2004)	Hedonic pricing	DK	-	layers	free-range eggs	-	per kg	15
Bennett (1997)	Contingent valuation	GB	591	layers	free-range eggs	0.43	per dozen	30.71
Bennett and Blaney (2002)	Contingent valuation	GB	164 (students)	pigs	legislation	1.09	per week	5.02
Bennett et al. (2002)	Contingent valuation	UK	119 (students)	slaughter animals	legislation	1.60	per week	5.75
				layers	legislation	0.94	per week	3.38
Bennett and Blaney (2003)	Contingent valuation	UK	591	layers	legislation	0.43	per dozen	-
				farm animals	animal welfare	5.50	per week	-
Bennett and Larson (1996)	Contingent valuation	USA	137 (students)	layers	free-range eggs	0.23	per dozen	18
				layers	legislation	5.13	per 1 year	-
				veal	legislation	5.12	per 1 year	-
Burgess and Hutchinson (2005), Burgess et al. (2003)	Contingent valuation	NI	192	layers	legislation	2.95	per week	-
			191	dairy cows	legislation	2.89	per week	-
			192	broilers	legislation	2.63	per week	-
			192	pigs	legislation	2.10	per week	-
Carlsson et al. (2005a)	Choice experiment	SE	710	broilers	chicken	1.31	per kg	-
				beef cattle	ground beef	0.50	per kg	-
				pigs	pork chop	0.57	per kg	-
				layers	egg	6.46	per dozen	-
Carlsson et al. (2005b)	Choice experiment	SE	827	pigs	chicken	-0.43	per kg	-
				beef cattle	ground beef	0.34	per kg	-
Carlsson et al. (2007a)	Choice experiment	SE	450	layers	free-range eggs	1.94	per dozen	-
					legislation	1.50	per dozen	-
Carlsson et al. (2007b)	Choice experiment	SE	757	broilers	chicken	-0.35	per kg	-
				beef cattle	minced beef	0.30	per kg	10
Dickinson et al. (2003)	Experimental auction	USA	54	pig	ham sandwich	0.31	per piece	16
				beef cattle	roast beef	0.39	per piece	20
		CA	54		sandwich	0.62	per piece	19
				pig	ham sandwich	0.42	per piece	13
				beef cattle	roast beef sandwich			
Glass et al. (2005)	Contingent valuation	I	1876	pigs	food	1.43	per week	-
Gracia et al. (2009)	Experimental auction	ES	73	pigs	cured ham	4.10	per kg	28

Kehlbacher et al. (2010)	Choice experiment	GB	282	beef cattle	beef	1.15	per year	6.52
			"	pigs	pork	0.81	per year	4.59
			"	broilers	chicken	0.97	per year	5.50
			291	beef cattle	legislation	26.30	per year	0.65
			"	pigs	legislation	19.20	per year	0.47
			"	broilers	legislation	17.80	per year	0.44
Kehlbacher et al. (2010)	Contingent valuation	GB	278	beef cattle, pigs, broilers	meat	4.30	per year	25.26
			300	beef cattle, pigs, broilers	legislation	121.72	per year	2.99
Lagerkvist et al. (2006)	Choice experiment	SE	285	pigs	pork	6.39	per kg	95.33
Liljenstolpe (2008)	Choice experiment	SE	1250	pigs	pork fillet	1.81	per lb	9.25
Lusk and Fox (2002)	Contingent valuation	USA	648	beef cattle	beef	-		17
Lusk et al. (2003)	Choice experiment	F	93	cattle	ribeye steak	6.46	per piece	-
		D	45			4.74	per piece	-
		UK	109			4.80	per piece	-
		US	566			5.28	per piece	-
Lusk et al. (2007)	Choice experiment	USA	594	pigs	pork chops	1.22	per kg	19
Moran and McVittie (2008)	Contingent valuation	UK	318	broilers	legislation	73.53	per year	-
Morkbak et al. (2010)	Choice experiment	DK	1322	pigs	minced pork	1.94	per kg	34.64
Napolitano et al. (2008)	Experimental auction	IT	104	dairy cows	yoghurt	0.16	per piece	
Napolitano (2009)	Experimental auction	IT	150	beef cattle	organic beef	8.60	per kg	49.76
Nocella et al. (2010)	Contingent valuation	IT,D, GB,ES,F	1416	farm animals	meat, dairy products, eggs, cheese	9.37	per week	-
Norwood and Lusk (2008)	Calibrated auction-conjoint valuation	USA	291	layers	eggs	0.60	per dozen	-
				pigs	pork chops	1.44	per kg	-
Olesen et al. (2010)	Real choice experiment	NO	115	salmon	salmon fillet	1.67	per kg	15
Rolfe (1999)	Contingent valuation	AU	100	layers	free-range eggs	2.83	per year	-
Taylor and Signal (2009)	Contingent valuation	AU	1224	layers, dairy cows, meat animals	eggs, dairy, meat	-		5-10
Tonsor et al. (2009)	Choice experiment	USA	205	pigs	pork chop	3.02	per kg	-
					legislation	0.49	per kg	-

*conversion rate of year of publication
Developed further from: Lagerkvist and Hess, 2011.

Valuing farm animal welfare by asking respondents to state their preferences for changes in husbandry practices has three shortcomings. First, every time technology and science find new welfare improved husbandry techniques, or make the existing ones less expensive, the valuation study has to be done afresh (McInerney, 1994). In addition, perceptions of animal suffering are influenced by what people consider to be acceptable or unacceptable which, in turn, depends on the perceived availability of alternatives. Thus, if alternative husbandry practices are being developed, people's attitudes and perceptions of existing husbandry practices and thus society's valuation of farm animal welfare change (Bennett, 1995). Second, laypersons who have been given a short introduction to the valuation method and the welfare issues at hand may have preferences for a certain kind of husbandry practices. However, they do not know what the changes in husbandry practices actually imply in terms of the welfare of the animals. Their valuation thus depends on how they perceive the husbandry practices to affect animal welfare, which is going to be different for different individuals. Third, the technical nature of the husbandry practices is likely to pose cognitive difficulties in a valuation study, in particular in choice experiments. These are prone to this kind of problem due to their necessarily repetitive nature and respondents having to evaluate relatively large and complex choice sets (DeShazo and Fermo, 2002). For example: respondents were asked to ponder whether they prefer pigs to be 'reared in a pen with 8 pigs (size = 0.90 m2 / 100 kg pig) or 'reared in deep litter with 50 pigs (size = 1.3 m2 / 100 kg pig)' (Liljenstolpe, 2008); broilers to be 'always kept indoors' or 'kept outdoors in summertime and in smaller groups' (Carlsson *et al*., 2007a); or to have a stocking density of 38 kg / m2 or of 34 kg / m2 and to experience periods of darkness for '8 hours with at least 4 hours continuous' or '8 hours continuously' (McVittie *et al*., 2006). These examples highlight the difficulties that respondents are faced with when stating their preferences. Arguably, even animal welfare scientists or veterinarians would have problems making these kind of decisions. Hence, value estimates obtained in this way from lay persons are likely to be somewhat unreliable, in a behavioural sense, and their validity may not be very high.

Instead of respondents having to figure out the welfare implications of different husbandry practices, they could be asked to state their preferences for animals experiencing a certain welfare state or level. This approach would require animal welfare scientists to carry out beforehand an impact assessment of different husbandry practices for the species under consideration on animal welfare. A study akin to this was carried out by Bennett *et al*. (2004) who asked experts to assess how UK policies, farm assurance schemes, and retailer specifications impacted on farm animal welfare. The experts' answers were then expressed in the form of numerical scores. This kind of welfare score could also be used in a farm animal welfare valuation study. It would alleviate the cognitive burden of the valuation task and, therefore, increase the study's reliability and validity. In addition, using a welfare score in the valuation study and obtaining a value for changes in the score could help policy makers and industry actors to better allocate resources to animal welfare which best improve welfare scores and are most valued by society.

A rather similar valuation study using a welfare score or measure has been carried out by Kehlbacher (2010). The score was based on theWelfare Quality®Index, an index developed by the European Commission-funded Welfare Quality®Project to form the basis of a European standard for evaluating the welfare of cattle, pigs and poultry in farms, and during transport and slaughter (Botreau *et al*., 2007a). Being applicable to different species, as is the Welfare Quality®Index, a

welfare score can be used in conjunction with the choice experiment method to obtain information about how the welfare improvement of one species may compensate for a worsening of the welfare of another species. Since preferences for farm animal welfare appear to be animal specific (Carlsson *et al*., 2005a), this kind of information could help policy makers determine how to achieve an overall level of animal welfare benefits in the most cost-efficient way. Finally, a generally accepted welfare score like the Welfare Quality®Index could form the basis of future farm animal welfare valuation studies so as to allow the comparison of valuation estimates between different species, production systems and countries.

Presenting farm animal welfare in the form of a score implies that animal welfare can be conceptualised as being measurable on a cardinal scale. Cardinality is a useful concept for the purpose of welfare economics because it allows the inclusion of farm animal welfare into the human utility function, thus making it a subset of human welfare. Moreover, cardinality ensures that inter-animal, inter-species and inter-systems comparisons can be made, and it allows the measurement and valuation of incremental welfare changes resulting from individual interventions as well as broad changes to states of animal welfare across the whole livestock production sector. The latter, in turn, eliminates the need for separate, repeated valuation exercises for every change in policy or animal husbandry practice that may have an impact on animal welfare. This has substantial resource implications. However, a valuation study using a cardinal welfare score may suffer from respondents' inconsistencies of understanding and misperceptions of the score and the current welfare state of the animals, which serves as a reference point. In addition, despite the animal science involved in the development of the welfare score, animal welfare is ultimately a human perception (McInerney, 2004), and the choice of methodology, scientific measures, and their interpretation are all based on subjective value judgements (Bennett *et al*., 2002). It is on this basis that the welfare score maps the welfare states of the animals into a single score, and the subjective nature of the welfare score clearly has implications for the conclusions drawn from any study using this approach.

Valuation of animal welfare according to a welfare score, rather than valuation of specific changes to livestock production practices, provides greater opportunity for benefits transfer of valuations on the assumption that the same improvement in welfare score is worth the same to citizens/society regardless of how that improvement is brought about. In contrast, the valuations of previous animal welfare valuation studies are generally not transferable because they each relate to very specific changes in animal husbandry/production practices.

6. The policy evaluation tool

6.1 Functional requirements of the policy evaluation tool

Defra has identified in the research call documents a set of requirements for a tool to aid the evaluation of welfare policies, which they described as a 'valuation framework'. These requirements collectively map out the 'work' that the tool is required to do. If the tool developed by this project is to be based on existing welfare valuation systems or frameworks, therefore, these must be evaluated in light of these requirements. To facilitate this, a set of evaluation criteria are needed, based on the key requirements set out for the policy evaluation tool. Table 5 below lists all of the requirements of the policy evaluation tool as specified in the project call and the contract document. These requirements have been given a thematic classification, and their role in defining the welfare measurement and monetary valuation components of the policy evaluation tool is separately identified.

Based on this requirement set, a set of evaluation criteria can be identified which capture all the key issues. At this stage, evaluation criteria will be drawn up for the 'welfare measurement' component of the policy evaluation tool only (Table 6). The requirements for the societal valuation component of the tool will be explored in Chapter 7. Because there is some duplication within the 'welfare measurement' requirements, rather than construct a criterion for each of the requirements listed above, a criterion will be constructed for each of the unique 'messages' contained within the set.

In order to arrive at a final set of evaluation criteria, a number of procedural steps have to be followed. First, it will be necessary to identify omissions (Step 1), i.e. requirements or constraints on the policy evaluation tool that are not contained in the list below. If such are detected, new evaluation criteria will be added. Following this, each evaluation criteria will be reviewed so that its purpose and function is fully understood (Step 2). If any problems with the evaluation criteria (and the 'requirements' that they represent) are revealed at this stage, criteria can be amended or dropped as required. The purpose of Step 2 is to determine which of the evaluation criteria are key to the functioning of the policy evaluation tool and have to be retained and which should be dropped either due to their peripheral nature, or because of methodological problems in applying them. Following this, the evaluation criteria will be ranked in terms of their importance, prior to their use in an Multi-Criteria Analysis (MCA) exercise to evaluate existing welfare measurement systems in terms of their usefulness as a basis for the new policy evaluation tool. The need for a ranking of the evaluation criteria and the creation of a subset of criteria for use in the MCA is not at first apparent. However, given a little thought it is obvious that no single existing framework for measuring animal welfare will meet all of the requirements set out below. It is also entirely possible that even a purpose-build policy analysis tool would struggle to meet all requirements. For this reason we need a means to decide which of the deficiencies of the existing systems are critical and which are not. The ranking of the evaluation criteria does just this.

Table 5. Functional requirement for the whole valuation framework

		Area of impact	
Theme	**Requirements**	**Welfare measurement**	**Societal valuation**
Scope	Should include measurement of marginal (incremental) welfare changes resulting from individual interventions for individual species.	✓	
	Should include measurement of broad changes to states of animal welfare across the whole livestock sector.	✓	
	Should measure the welfare status of farm animals on a 'continuous' welfare scale (necessary to capture fine changes in welfare states).	✓	
	Should provide benefit valuation 'referents' that are capable of transfer across policies and husbandry methods.	✓	✓
	Should indicate whether or not improvement is achievable and measurable.	✓	
	Make societal benefits of welfare changes apparent.		✓
	Provide a benefit valuation that can be used within a cost-benefit framework (i.e. can be monetised).		✓
Methodology and information base	Requires robust, information-based methodology for measuring animal welfare.	✓	
	Must capture in a transparent way the technical relationship between animal production practices and states of animal welfare.	✓	
	Should measure the welfare status of animals by a combination of resource/input and welfare outcome criteria.	✓	
	Should be based on sound scientific research, practical experience and other relevant evidence, derived from both animal/veterinary science and social science.	✓	✓
Data outputs	The welfare benefits associated with government interventions, accounting for: • The number of animals affected. • The duration of the welfare problem experienced by each affected animal. • The impact of the problem on, each of the Five Freedoms.	✓	
	Account for the average level of welfare enjoyed by livestock and any episodes of welfare above or below the average	✓	
	Valuation of societal benefits		✓
Modernity and Shelf-life	Permit regular up-dating, at relatively low cost, as and when new animal welfare valuation data become available.	✓	✓
	Reflect recent advances in animal welfare science.	✓	
	Contain the very latest data on welfare standards and welfare benefits.	✓	✓
	Contain the very latest data on societal valuation of welfare benefits.		✓

At the conclusion of this process, a detailed understanding of the requirements of the welfare measurement component of the policy evaluation tool will be available. This understanding will facilitate the identification of problems likely to be encountered in constructing the evaluation tool and, hopefully, strategies to overcome them.

Once the a set of evaluation criteria have been derived, these will be used as the basis of an MCA, which will be applied to the existing metrics and frameworks for measuring the welfare states of kept animals. The outcome of the MCA exercise will be a transparent, methodical, and replicable critical evaluation of existing approaches to welfare measurement in terms of their suitability for use in the current framework.

Table 6. List of evaluation criteria based on 'requirements' above

Scope

- Suitable for measurement of welfare for individual species
- Provide a metric (continuous scale) sensitive enough to capture small-scale incremental changes in welfare
- Provide common metric suitable for measuring welfare across species
- Provide a metric which 'interprets' changes in policies and husbandry methods in terms of welfare changes.
- Indicates whether improvement is achievable and measurable.

Methodology and information base

- Evidence-based methodology for measuring animal welfare (from both animal/veterinary science and social science).
- Metric based on technical relationship between animal production practices and states of animal welfare.
- Methodology used to derive technical coefficients is scientifically robust and transparent.
- Welfare metric should be based on both resource/input and welfare outcome measures.

Data outputs

- The welfare metric accounts for the number of animals affected.
- Account for the average level of welfare enjoyed by livestock as well as / any episodes of welfare above or below the average..
- The welfare metric accounts for the duration of welfare problems experienced
- The welfare metric accounts for the impact of the welfare problem on the Five Freedoms.

Modernity and shelf-life

- Amenable to regular and low cost up-dating.
- Reflects recent advances in animal welfare science.
- Contains the very latest data on welfare standards.

Step 1 – identification of omissions
No omissions have been revealed.

Step 2 – review of individual evaluation criteria
In the sections that follow each of the proposed evaluation criteria is reviewed in turn. The review in each case covers four issues:

- an elaboration of the requirement, or constraint, underlying the criterion
- the relevance of the criterion to the welfare measurement component of the policy evaluation tool
- any problems or difficulties that exist with the criterion as a construct, or with its application
- the importance of satisfying the criterion to the functioning of the welfare measurement component of the policy evaluation tool. Importance is rated on a three point scale, i.e.
    - low importance – the tool could still function well with this requirement unfulfilled;
    - moderate importance – the tool would function to some extent with this requirement unfulfilled, but key areas of functionality would be lost; and
    - high importance - the requirement is considered critical to the functioning of the tool and without it the tool would not function in any meaningful way.

6.2 The evaluation criteria

- Suitable for measurement of welfare for individual species

It is a requirement of the policy evaluation tool that it have the facility to measure the welfare state of individual species and this evaluation criterion encapsulates that requirement. The requirement is essential to the point that some might think it hardly worth defining in this formal way, i.e. as an evaluation criterion. It is proposed to include it for completeness sake.
Importance rating – HIGH.

- Provide a metric (continuous scale) sensitive enough to capture small-scale incremental changes in welfare

This evaluation criterion envisions the possibility that Government may want to evaluate the welfare effect of given policy/practice changes that may have small-scale gross effects on welfare states, but where these changes might be greatly valued by society.

There would be several problems and limitations with a welfare metric that operates at such a fine level. First, minor changes to welfare states might be very difficult for society to understand and value. Also, the smaller the signal being generated, the greater the chance that measured 'change' is just 'noise'. A further potential problem is one of spurious accuracy, i.e. a highly 'sensitive' measurement system used within species would be pointless, if aggregation over species is required. For example, suppose a new broad-brush policy measure is introduced and this is estimated, using a welfare measurement tool which employs a 100-point metric, to lift the welfare score of sheep by 1 point, dairy cows by 6, and pigs by 10 etc. A single score for the whole livestock sector would have to take some form of mean of these change values (possibly a

weighted mean) and by so doing the estimated mean value would be only 'approximately' representative of individual species. To maintain the value of the fine-scaled metric the framework would have to be limited to generating a suite of metrics, i.e. one for each species.
Importance rating – MODERATE (on the grounds that the welfare measurement component of the policy evaluation tool would be quite possible and functional operating at coarser scales).

- Provide common metric suitable for measuring welfare across species

This requirement is driven by a desire to provide a simple unified metric of welfare covering the whole of the livestock sector. This anticipates the introduction of policy or regulation that impacts on several types of livestock simultaneously, where a common metric would allow comparison of the impacts of that policy on multiple livestock sectors. The use of the common metric does not require that a single welfare score be produced (see above) averaging across all livestock sectors, although this would, in theory, be possible. Neither does the use of a common metric require that the technical coefficients used to translate changes in management practice into changes in welfare state be identical across livestock sectors. It simply means that the welfare 'scores' for each livestock sector are derived using a common measurement system.
Importance rating – HIGH.

- Provide a metric which 'interprets' changes in policies and husbandry methods in terms of welfare changes.

This evaluation criterion duplicates the criterion under the 'Methodology and information base' theme, i.e. 'Metric based on technical relationship between animal production practices and states of animal welfare'. For reasons of efficiency, this duplicate criterion should be dropped.

- Indicates whether improvement is achievable and measurable.

The policy evaluation tool is required to measure states of animal welfare and provide an indication of societal valuation of these changes. The tool, depending on its exact specifications, will have a given level of maximum resolution. It should be a simple matter therefore to indicate whether a given set of proposed changes to practice will result in a 'measureable' change in welfare state. This does not preclude the possibility that actual changes in welfare state might have occurred, only that these are small to be reliably measured using the chosen metric.

What such a tool cannot do, is assess the extent to which particular target changes in welfare state are achievable. For example, Government might propose a set of changes to livestock management practice and the framework could be used to indicate likely changes to welfare state that might result from this, assuming that the changes have sufficient magnitude to be measurable. However, this would not guarantee that these changes to welfare state would be achievable in practice, because other factors may intervene to prevent this, such as industry avoidance of the management changes, perhaps on cost grounds. The changes in welfare state that the policy evaluation tool projects (i.e. identified with particular changes in management practice) should therefore only be seen as indicative, not definitive. Over time, with repeated use of the tool, these estimates can be 'calibrated' by comparison of projections with the actuality. What any tool can be expected to do is indicate, in crude terms, the maximum level of welfare

improvement that is obtainable by the optimal mix of management and resources. However, all available welfare measurement tools would supply such estimates, with greater or lesser reliability.

Because any policy evaluation tool, no matter how constructed, will have limits to its measurement capability and these limits can be known, an evaluation criterion based on the ability to indicate 'measurability' does not distinguish between approaches. Because no tool can indicate with accuracy how achievable, in practice, welfare improvements are, this evaluation criterion also does not distinguish between approaches. In view of the above, this evaluation criterion should be dropped.

- Evidence-based methodology for measuring animal welfare (from both animal/veterinary science and social science).
- Metric based on technical relationship between animal production practices and states of animal welfare.

It is a requirement of the project that the policy evaluation tool is based as far as is possible on scientifically–derived evidence. The evidence base would need to underpin two aspects of the tool, (i) the choice of the measurement metric; and (ii) the coefficients linking management practice (including resources/inputs) to welfare states. Ideally, (ii) would be derived from field trials of production environments and management practices and direct observation of associated animal welfare states. The correlation between the two would ideally be estimated using statistical methods. However, it is accepted that much of the data that will be available to construct the welfare measurement framework will have been derived by methodologies that fall short of this ideal. It is likely that some types of data will be in short supply and pragmatic choices will need to be made about what data is of an acceptable quality (reliability and modernity etc) and can be used and what should not. For example, it is possible that some of the data available will be in the form of expert opinion, and a judgement will have to be made concerning whether this is better than no data at all. Obviously, the older the existing data, or evaluation framework, the less likely it is that it will reflect the evidence of the latest empirical studies.
Importance rating (in order of presentation) – MODERATE & HIGH respectively

- Methodology used to derive technical coefficients is scientifically robust and transparent.

There are two aspects to this criterion. First, if technical coefficients are derived from primary data specifically for the construction of this policy evaluation tool, the methodology used in their estimation must be both rigorous and transparent. Second, if existing tertiary data, derived from past studies, are to be used in the construction of the tool, the methodologies by which the data were derived should be both transparent and scientifically robust. The data likely to be most difficult to assess in this way will be that derived from expert opinion, as the referents and assumptions used in these cases are almost always unknown. This is not to say that data derived by these means should not be included in such a tool – such approaches are commonly used in the absence of suitable empirical evidence, to plug gaps in data, or where there is no systematic framework to allow the distillation of diverse data sources. Such data can be of considerable value, if reliable processes have been used to derive them, for example through use of MCA or Delphi techniques. These techniques have been used to great effect in the area of human health

valuation, for example, in the construction of DALY and QALYs, where expert groups have been used to rank order various sequelae in terms of the severity of their impacts on quality of life.
Importance rating - HIGH

- Welfare metric should be based on both resource/input and welfare outcome measures.

In the past, welfare measurement frameworks have been largely based on assessment of the input or resource side of production systems (i.e. the environment in which the animal is kept). The focus on this type of 'indicator' has been largely driven by their low cost and ease of measurement, i.e. such measures can be constructed largely from paper records. Under this approach, certain assumptions are made about the way in which the welfare of animals is affected by changing the physical environment in which they are housed, and other management practices. As the literature review has shown, in recent years there have been moves to include direct observations of health states in welfare measurement systems and the very latest welfare measurement approaches have extended this from observation of purely physical condition to include animal behaviours. Few would argue that these changes have made welfare measurement systems more detailed and comprehensive, but in the process the use of such measures has ceased to be desk-based. The inclusion of outcome-based measures of welfare presents an obstacle to using, for example the Welfare Quality approach, as a template for the policy evaluation tool proposed in the current project. It is a requirement of the proposed tool that it be capable of translating policy proposals into changes in welfare states. Within that frame, it may not be possible to carry out on farm-observation of health and welfare outcomes, because the policies have not been implemented and therefore have had no effect on welfare outcomes. In order to use a modern welfare measurement system such as Welfare Quality as the basis for the proposed tool, the requirement for observational measurement may have to be relaxed.
Importance rating – HIGH.

- The welfare metric accounts for the number of animals affected.
- Account for the average level of welfare enjoyed by livestock.

Existing methodologies for measuring animal welfare do two things, they either provide a measure of average welfare state for a given population (a measure of central tendency), or they provide the proportions of the population in particular welfare classes. Some do both. These approaches do not directly account for the total number of animals in different welfare states, although some account is taken of this indirectly, in defining the population from which the estimates are generated – this might be the animals on a particular farm, or in a region. The problem with generating a single measure of central tendency for a species is that it reveals nothing about the distribution of the sample and, in particular, the extent of extreme values. This missing information is fairly crucial, because it is at the extremes that the worst welfare problems are found and where most remedial effort should be directed. In order to properly account for the number of animals affected (i.e. in particular welfare states), the proposed tool will either have to apply some form of weighting process in deriving a single measure of central tendency for a species, or the metric will have to reflect the distribution over levels of welfare state, as well as define the size of the total population. In terms of weighting for animal numbers, while it is important that the policy evaluation tool proposed by the study is able to do this do this, existing

evaluated measurement systems need not be discounted if they do not do it, because it may be an easy matter to adapt them to do so.
Importance rating (in order of presentation) – MODERATE & LOW.

- The welfare metric accounts for the duration of welfare problems experienced.

While the criterion above is framed in terms of welfare problems, the policy evaluation tool is framed in terms of welfare gains. The two are obviously inter-related, but for consistency sake, further reference to this issue will be framed in terms of the duration of welfare gains. This evaluation criterion, which appears in the Call document, is taken directly from the Defra Animal Welfare Delivery Strategy (2007). The issue of the duration of welfare gains has not been addressed by existing welfare measurement systems, which simply record welfare state at the point (in time) of measurement. In the context of the proposed policy evaluation tool, is it reasonable to expect that it be able to account for the duration of welfare gains for the various species? The duration of welfare improvements will depend on: (i) the life span of the animal; and (ii) the point in the life cycle that the intervention is applied. The possibility of capturing the effect of remaining life-span in the proposed welfare measure is somewhat limited. This limitation would be less constraining if the tool were being designed for on-farm welfare measurement, where a welfare score could be collected for each individual animal and its age at the time of intervention were known. However, this is not the purpose of the tool. This is not to say that nothing can be assumed about the lifespan of animals beyond the point of intervention, as particular interventions can be associated with particular parts of the production cycle. For example, one intervention might apply to beef animals from the day of weaning, while another might be applicable during transport to slaughter. The first would impact welfare by a given amount for, say, 16 to 22 months, while the latter would impact welfare for a matter of hours. The question then arises, how are these to be reflected in a common metric? How much does reduced stress over a few hours before slaughter equate in welfare terms to, for example, better foot condition over a lifetime? This issue will be subject to further discussion below.
Importance rating – HIGH.

- Account for any episodes of welfare above or below the average.

This evaluation criterion (and the requirement upon which it is based) is ambiguous. It could be taken to mean either: (i) the enumeration of frequency of occurrence of welfare states in the population which are higher and lower than the population average; or, (ii) recognition of variation in the welfare state of individual animals over their lifespan, where this variation is referenced to the population mean. The first of these requirements is possible and has been dealt with above. The second is not feasible on the grounds that: (a) the tool is not designed to measure the welfare of individual animals; and (b) the measurement of welfare over whole populations of animals reflects an average over animals at different stages in their lifecycle, and at this level of aggregation the distinctions between lifecycle stages are not apparent. On the basis of this assessment, this evaluation criterion should be dropped.

- The welfare metric accounts for the impact of (welfare) problems on the Five Freedoms.

This evaluation criterion, which appears in the Call document, is taken directly from the Defra Animal Welfare Delivery Strategy (2007). This evaluation criterion (or the requirement from which it is derived) effectively requires that the policy evaluation tool be based on the Five Freedoms system. The Farm Animal Welfare Council's Five Freedoms (FAWC, 2011) are a set of aspirational targets for animal welfare, rather than a protocol or tool for measuring welfare. For this reason this evaluation criterion is not relevant for the construction of the welfare measurement tool itself, but it will be used to guide the type of metric that the measurement system employs. The metric will thus have to be 'interpretable' in terms of the 5 Freedoms, at the very least indirectly, as these now permeate much of the policy and welfare assessment in this country, the EU and internationally.

- Reflect recent advances in animal welfare science.
- Contains the very latest data on welfare standards.
- Amenable to regular and low cost up-dating.

While the policy evaluation tool (or rather the welfare measurement component of the tool) that is developed in light of this project might employ the most up-to-date data on welfare standards and approaches to welfare measurement, these requirements acknowledge that the landscape of animal welfare measurement is rapidly changing and that periodically, the tool that is developed in light of this study will need to be updated, to reflect new policies and standards, and to make use of new data arising from academic and other research, as well as industry practice. In terms of designing a methodology for the policy evaluation tool (as distinct from the data that populate the tool), it is quite possible that an older welfare measurement system might prove the most appropriate for the current requirements. For this reason, it is not essential that reviewed approaches contain the very latest scientific data, as it may be possible to add these from elsewhere.
Importance rating (in order of presentation) – MODERATE, MODERATE, HIGH.

6.3 Critical evaluation of existing welfare measurement systems

In the MCA exercise that follows a number of existing welfare measurement systems are critically evaluated on the basis of the evaluation criteria that have been identified and refined in the discussion above. Three existing welfare measurement systems are evaluated: the German ANI 200 system, the Austrian ANI 35 L and the Welfare Quality Protocol. For completeness, the proposed policy evaluation tool, or rather, an informed theoretical understanding of its necessary properties, is evaluated alongside these. While these properties will not be apparent to the reader at this juncture, these will be explained in the sections that follow.

Before moving on to the critical evaluation itself, a word of explanation is perhaps warranted on the choice of only these three current systems for appraisal (for consideration as a possible basis for the proposed policy evaluation tool) and also the reason for the focus on 'complete' welfare measurement systems to the exclusion of the wealth of other data in the literature relating to the issue of animal health and welfare.

Two types of study are available in the literature that have potential to contribute data in support of the welfare measurement component of the new policy evaluation tool:

(i) innumerable empirical studies that have attempted to identify the risk factors associated with animal health problems, or welfare improvements; and

(ii) complete welfare measurement systems, which by various means have encapsulated the evidences derived from (i).

The studies represented at (i) are very diverse, both in terms of their objectives and their methodology. Some seek to understand the environmental or management causes of health problems through observation (measurement) and tests of statistical association, while others make direct interventions (changes to environment or management) to monitor outcomes. Some are interested in the influence of individual environmental (incl. management) factors, while others are interested in baskets of factors (sometimes whole production systems). Some have as a focus individual health conditions, such a pododermatitis or salmonella infection, while others are interested in general health. Some studies consider outcomes solely in terms of physical health, others psychological state, others more generalised measures of welfare. Some studies are carried out within a single tightly defined production system, while others transcend production systems, or test new systems. The great majority of studies are carried out for individual species, and often individual breeds within species. This catalogue of bewildering diversity serves to illustrate the extreme difficulty that would attach to any attempt at extracting data from such studies for use as technical coefficients in a welfare measurement system. It is for this reason that past large-scale attempts to construct welfare measurement systems have had to rely on expert systems to process and distil this data and it is for this reason that we look to these expert systems as the source of existing data capture for use in a new policy evaluation tool.

The choice of these three existing welfare measurement systems for the evaluation process is driven by the fact that, of the very limited number of such systems in existence, these three are still in common use for practical measurement of animal welfare.

Each of the evaluation criteria defined above has been weighted on a four-point scale reflecting the importance of each underlying requirement (or characteristic) of the welfare measurement component of the new policy evaluation tool. Each of the three existing welfare measurement systems is scored on the basis of each evaluation criterion (Table 7). To simplify matters, the measurements systems are scored using an ordinal scale of: double negative (–), one negative (-), one plus (+) and double plus (++). A negative score on any of the criteria ranked as being of high importance indicates that the measurement system in its current state fails in terms of an important requirement of the new tool. Where the rank score is replaced by a question mark, this signifies that not enough is known about the system under evaluation to make a particular judgement. These evaluation judgements have been undertaken by the members of the project team.

Table 7. A critical evaluation of three existing animal welfare measurement systems and the proposed policy evaluation tool

	Evaluation criteria												
	Suitable for measurement of welfare for individual species	Provide a metric (continuous scale) sensitive enough to capture small-scale incremental changes in welfare	Provide common metric suitable for measuring welfare across species	Evidence-based methodology for measuring animal welfare (from both animal/veterinary science and social science).	Metric based on technical relationship between animal production practices and states of animal welfare.	Methodology used to derive technical coefficients is scientifically robust and transparent.	Welfare metric should be based on resource/input measures.	The welfare metric accounts for the number of animals affected.	Account for the average level of welfare enjoyed by livestock.	The welfare metric accounts for the duration of Improvements to welfare state.	Reflect recent advances in animal welfare science.	Contains the very latest data on welfare standards.	Amenable to regular and low cost up-dating.
Importance rating	High	Mod	High	Mod	High	High	High	Mod	Low	High	Mod	Mod	High
Welfare Quality	++	-	+	+	-	++	+	-	+	-	++	?	+
ANI 35 L (Austrian)	++	+	+	+	+	+	++	-	+	-	-	-	+
ANI 200	++	+	+	+	+	?	++	-	+	-	-	-	+
The new policy evaluation tool	++	++	+	+	+	+	++	?	++	?	++	++	++

6.4 Review of suitability of existing welfare measurement systems

Provide a metric (continuous scale) sensitive enough to capture small-scale incremental changes in welfare

None of the current systems of welfare measurement perform particularly well on this criterion, primarily because of the level of aggregation at which they choose to present their final welfare scores. While the ANI 35L and ANI 200 systems provide a single welfare score, this is based on an ordinal scale with 6 categories, ranging from 'Not suitable' to 'very suitable'. However, underpinning this classification is a pseudo-cardinal (or ordinal) scale ranging from -11.5 to 45 which, if used, would make the measurement system much more sensitive to small-scale changes in welfare. The WQ system does not provide a single metric of welfare based on a cardinal scale, providing instead a single welfare outcome, derived by decision rules from four separate metrics reflecting, for example, the 'Principles' of good feeding , good housing , etc, where each of these consist of a 4-point ordinal scale. Underlying these classes is a percentage-scaled metric offering more detailed measurement, but potentially limiting here is the lack of a ready means to combine these four 'Principles' measures into a single cardinal welfare score. The authors' of the system assert that to combine the four Principles mathematically would break the rule that deficiencies on one dimension (Principle) cannot be compensated for by strengths on another dimension. Ways of overcoming this limitation in the WQ approach are discussed in later sections.

Provide common metric suitable for measuring welfare across species

All of the systems evaluated satisfy this criterion, i.e. there is commonality between species in terms of the metric and calculation methodology used. Obviously there must be differences in the observational measurements taken and technical coefficients used to link risk factors (environment) to welfare states, as a consequence of differences in the physical and behavioural natures of species and the different rearing environments that these require.

Evidence-based methodology for measuring animal welfare (from both animal/veterinary science and social science).

All of the existing systems of welfare measurement and the proposed system perform reasonably well on this evaluation criterion as they all indirectly take into account the body of scientific knowledge relevant to the issue at the time of their creation. However, none of the existing approaches score as highly as they might, as the empirical underpinning of the systems is not used in a transparent way, but rather is filtered through the medium of expert opinion (a necessary requirement when dealing with such a diverse body of evidence). On this basis the evaluation criterion is satisfied in all cases but no distinction between approaches can be made.

Metric based on technical relationship between animal production practices and states of animal welfare.

None of the approaches scores highly on this criterion because the relations between production environment and animal welfare are based on expert judgements rather than on

technical coefficients derived directly from empirical studies in a transparent way. The WQ system is penalised here in addition because its welfare measures require direct observations of health and mental state as well as measurement of environment and management factors. This focus on direct observation is a real benefit in a farm-based welfare measurement system designed to be applied to individual animals, but a handicap in a desk-based, policy-support system designed to generalise about whole populations.

Methodology used to derive technical coefficients is scientifically robust and transparent.
Accepting the points made above about the need to use expert judgements to derive technical coefficients in each of these systems, the manner in which these coefficients are obtained needs to be managed to maximise their robustness. Various multi-criterion decision techniques are used by the current systems to achieve this, with the most rigorous being those employed by the WQ project, although this does not increase transparency by any means. All systems perform satisfactorily on this criterion.

Welfare metric should be based on resource/input measures.
As discussed in the criteria selection section above, it would not be possible to include outcome-based (obtained by direct observation) factors in the construction of a desk-based welfare measurement system that meets the sponsor's requirements. On that basis, the WQ system performs poorly on this criterion, and adaptations to methodology would therefore have to be applied for this approach to be considered as the basis of the new policy evaluation tool. The ANI 35 L and ANI 200 systems perform well on this criterion because they rely solely on environment measures that do not necessarily require on-farm observation.

The welfare metric accounts for the number of animals affected.
None of the existing systems perform well on this criterion. Both the ANI 35 L, ANI 200 and WQ systems are designed to measure welfare of individual animals, and groups of animals, but are not designed to be applied to whole populations. The WQ system will provide estimates of the proportions of animals in small populations that fall into various welfare classes, but because of the system's very high data collection burden, it would be impractical to apply it to larger populations, although it would be possible to apply it to representative samples of animals/farms and from that to generalise to the whole population. The ANI35L system has much lighter data gathering requirements, in that most of the data required can be obtained from paper records. However, even this requires a farm survey of some kind and so the best that could be achieved would be the survey of a representative sample and generalisation from that. The new policy evaluation tool (as proposed below) could conceivably achieve this outcome by using expert judgement to determine how the welfare demographic of the animal population would change under a given set of changes to the rearing environment. However, without accurate knowledge of the pre-intervention (baseline) position (which can only be known by survey), there would be no grounding of these projections of change in reality.

Account for the average level of welfare enjoyed by livestock.
The AN I35 L, ANI 200 and WQ systems are not designed to generate a statistic such as the average welfare of livestock in a population. To begin with the WQ system does not in current practice provide a single cardinal-scale metric of welfare. A mean value would have to be taken from the percentage-based scale that underlies the four 'Principle' measures. However, while mean welfare values, of sorts, could be generated for all of the systems evaluated here, the usefulness of the statistic is debatable and it is for this reason that the weight attached to this criterion is low.

The welfare metric accounts for the duration of improvements to welfare state.
The existing welfare measurement systems do not do this. The opportunity exists to include such a function in a purpose built policy evaluation tool, but it is not easy to envision how this might be done, as there are a number of conceptual and practical difficulties that would first have to be resolved, such as the differing life-spans of producing animals and the requirement to value, within each species, a full life under various states of welfare, against a life cut short perhaps due to poor welfare, or even some event unrelated to welfare, such as predation. There will be more discussion of this issue, and the feasibility of the new tool delivering on it, in the sections following. For the present then, it is sufficient to acknowledge that all of the existing systems fail on this highly weighted criterion.

Reflect recent advances in animal welfare science.
The new policy evaluation tool proposed below, being the most recent, will obviously better reflect the latest advances in animal welfare science than the existing systems. However, in view of the relatively recent development of the WQ system, this should not offer much of an advantage. The data in the ANI 200 and ANI 35 L systems are known to be historic.

Contains the very latest data on welfare standards.
The three existing welfare measurement systems have no requirement to contain data on welfare standards, as their outputs are judged exogenously in relation to welfare standards, or are potentially used in creating these welfare standards. Welfare standards are not therefore embedded within them. The same is true of the new policy evaluation tool. Where welfare standards are expressed in terms of environment or management factors, all of the systems can measure or estimate (after their own fashion) the effect of these standards on welfare.

Conclusion
The critical evaluation above shows that none of the existing welfare measurement systems is wholly suited to meeting the requirements of the proposed policy evaluation tool. However, it is apparent that the WQ system is the approach that offers the most useful functionality and therefore this has been selected as the basis of the new tool, subject to a certain amount of adaptation, which will be explained in the sections following. The ANI 35 L system, and the earlier German ANI 200 system, have been rejected on the grounds that the technical coefficients contained in them were developed quite a long time ago and will therefore not

reflect recent advances in animal welfare science. None of the systems reviewed account for the duration of changes in the state of animal welfare, so this criterion is not discriminatory.
An apparent limitation with the WQ system is the lack of a single 'cardinal' scale in welfare measurement and this might constrain the use of the WQ system as the basis of the policy evaluation tool. However, it is apparent that underlying the four 'Principle' welfare measurements are cardinal scales that can be combined to yield a single aggregate measure. Indeed, the WQ methodology does yield a single welfare classification of farms, based on the application of a set of decision rules to the four 'Principle ' scales, rather than applying a simple mathematical aggregation. An adaptation of this aggregation process should therefore remove this limitation.

A more significant limitation of the WQ approach in the current context is the fact that the technical coefficients on which the system is based, are a mix of health-to-welfare and environment-to-welfare relationships, while in the policy analysis tool these two types of relationship need to be separated (the reason for this will be explained in the sections below on the design of the policy analysis tool). Adaptations to the WQ approach to overcome this limitation are also discussed in detail in the sections following, but in summary, there are two options: first, separate out the two types of coefficients and use these separately; and second, generate new sets of coefficients as needed, using the WQ indicators as a framework and possibly following the WQ methodologies for data generation.

A limitation of all of the existing systems, that has not been discussed thus far, is that none of them makes the link with policy, and so even using the WQ methodology as a basis, this link would have to be added. Policies are generally framed in terms of management and environment prescriptions and so new policies will likely change the state of these factors. This link would need to be based on expert judgements, which 'interpret' policy changes in terms of changes to environmental factors. This of course means that, even if no other adaptations of the WQ approach were necessary, the system could not be used 'off the shelf' for the purposes of policy evaluation, because it would be necessary to convene some kind of expert panel to pre-process the input (scenario) data.

A final limitation of the existing WQ system that will require some adaptation derives from the fact that the technical coefficients contained within the system are bounded by current, or historical, practice and conditions. This is not a big limitation if the only work that the system is asked to do is measure extant states of welfare, but it is a problem where the system is asked to project the impacts of changes to environmental conditions arising from new policies, because these changes might potentially place the rearing environment outside the range of historical experience. Without adaptation the WQ system would only be able to monitor the outcomes of these changes, but not project the changes in advance. The adaptations required to overcome this are discussed below, but in summary this adaptation involves a recognition that historic technical coefficients need to be updated in some way in the light of the provisions of new policy (if that new policy would establish rearing conditions that are beyond current experience).

6.5 A new policy evaluation tool

Process and product

The idea that a new policy evaluation tool can be static, operated by a single individual is erroneous. Perhaps the first thing that should be stressed about the properties of such a tool therefore is that it will be a process as much as a product. There are four parts to this process:-

(i) Because the best of the existing welfare measurement systems, the WQ system, does not fully meet the project requirements, an adapted system will have to be constructed, using some new technical coefficients (the type of coefficients will be discussed below). A process of construction of these technical coefficients will therefore have to be undertaken.

(ii) It is a key requirement that the new policy evaluation tool be amendable to regular and low-cost updating to take advantage of new scientific advances in the understanding of welfare issues. This process will involve periodic input from experts.

(iii) By nature, the introduction of new policies, or policy proposals, introduces a set of unknowns, i.e. the introduction of rearing environments that do not currently exist. Therefore, at the point of analysis of new policy proposals, an expert panel will have to be convened to generate a new set of coefficients linking policy provisions to environmental conditions. This process can be thought of as preparing scenario data.

(iv) The projections arrived at through use of the tool must be periodically validated, through comparison with counter-factual data i.e. empirical evidence from the real world. Discrepancies between projections and observed outcomes should form the basis of calibration exercises applied to the technical coefficients of the tool. In the normal course of events, this type of exercise will occur as a consequence of policy evaluations, where policies are evaluated to see that they are working well and giving rise to the expected benefits. Therefore, the results of such policy evaluation projects should yield the data required here.

For more on this issue, see the section entitled 'How would the policy analysis tool work in practice?' below.

Based on expert judgement

In common with the general methodology employed by existing welfare measurement systems, the technical coefficients linking (either directly or via intervening steps) policy provisions with states of animal welfare, will need to be derived from expert judgement. As has been suggested in the argumentation above (there is more on the reasons for this below), it may not be possible to 'lift' technical coefficients directly from the WQ system or from the animal welfare or health literature. The policy analysis tool will therefore have to be based on new expert judgements and will therefore form part of an expert system, which will to a greater or lesser extent be dynamic.

Best practice

Although it may not be possible to derive technical coefficients for the policy evaluation tool from the WQ system (or other extant sources), the methodologies employed in the WQ project in deriving such coefficients might be adopted where these are believed to represent best practice. In this regard, the WQ project represents a good exemplar, with its strong use of statistical techniques to support expert decision-making.

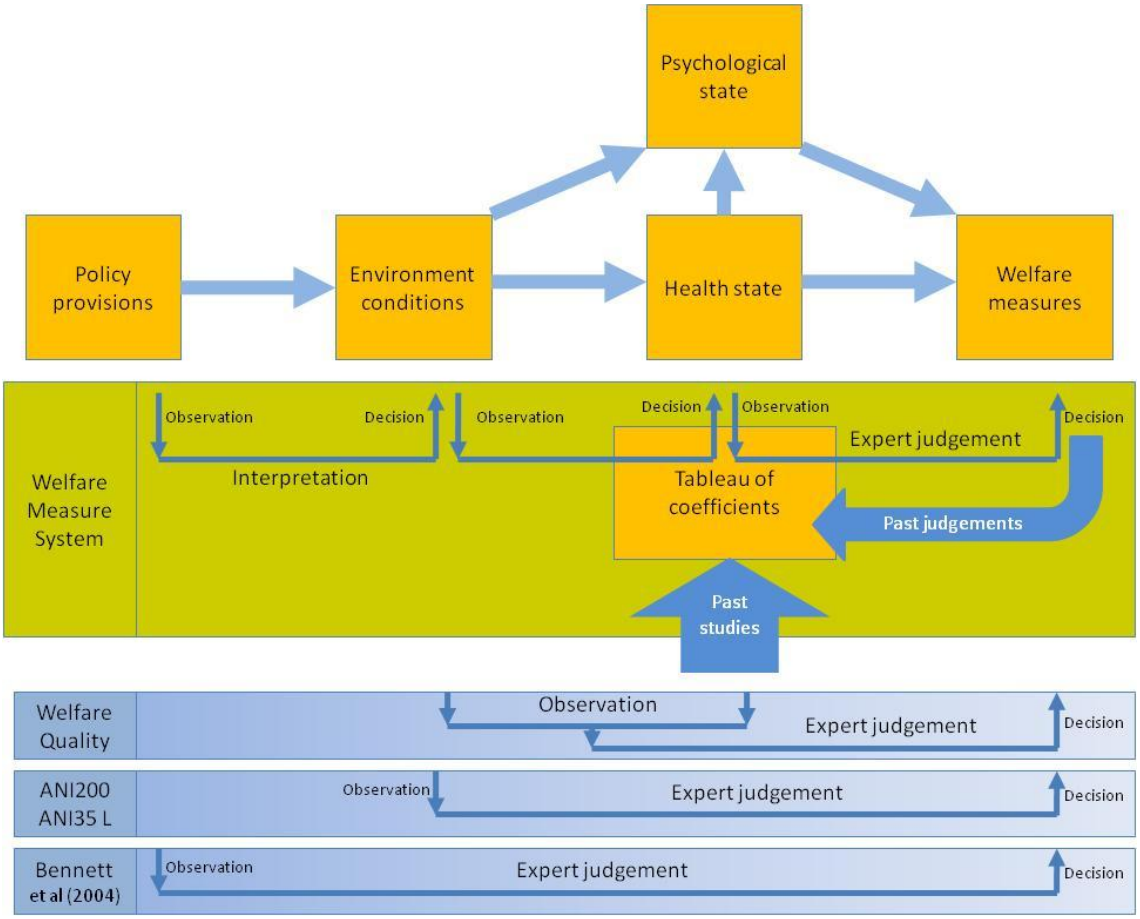
How do existing welfare measurement systems reflect the chain of causality between policy provisions and states of animal welfare?

Policy does not usually impact on animal welfare directly. Policy is usually defined in terms of physical environment and management prescriptions and as such impacts these things to produce a changed rearing environment. The rearing environment interacts with management competence to impact on the health and mental states of animals. Animal welfare, which is an artificial construct and a human perception, is variously defined (using some form of mathematical scale) on the basis of a combination of measures of the health and mental state of animals and also the rearing environment as illustrated in Figure 8 below.

Existing welfare measurement systems are based on coefficients that bridge the gaps between various parts of the chain of consequences outlined in the statement above. The technical coefficients of the ANI35L and ANI 200 systems bridge the gap between environment and welfare, while in the WQ system, the coefficients bridge the gap between, primarily, health outcomes and welfare state, as shown in the lower part of Figure 1.

The new policy evaluation tool (see Figure 1) will require three sets of coefficients, bridging the gaps between: (i) policy decisions and rearing environment; (ii) rearing environment and health state; and (iii) health state and the welfare measure. At this point, the question should be asked, why are three sets of coefficients necessary? Does this not simply mean that one set of expert judgments is being compounded with other sets of expert judgements, resulting in three sources of potential error, rather than one? Would it not be better to deploy a single set of judgements to bridge the whole chain? The short answer to this question is ‘no’. The longer answer follows.

Figure 1. Schematic showing chain of decisions/interpretation employed to arrive at measures of welfare state in past welfare measurement systems and the proposed policy evaluation tool.



(i) Linking policy provisions to environmental outcomes

The first set of coefficients would have to be generated whatever system was used and is therefore a necessary adaptation of the WQ system. Policy provisions are factors that affect the rearing environment and they have to be interpreted in those terms (i.e. in terms of the changes policies bring to the rearing environment). Even if a single 'bridging' set of judgements from policy to welfare state was attempted, the link between policy and rearing environment would still have to be implicitly explored and accounted for. Rather than have these judgements implicit, it would be better to make them explicit, so that their validity can be tested, perhaps most usefully by peer review.

Logic tells us that it would be relatively straightforward for experts to project the likely outcomes of policy provisions on the rearing environment, in part because the conceptual gulf is narrow, but primarily because policy provisions are themselves often expressed in terms of changes to environmental conditions, for example space allowances, lighting levels,

frequency of inspections etc. Because of this it might be expected that a fairly robust set of expert judgements on the changed rearing environment would be obtainable.

(ii) Linking environmental conditions to health outcomes

The proposed system adapts the WQ approach, by introducing a new step in the decision-making chain to capture the link between environment and health (and behaviour) states. The approach adopted in this case follows the approach of Bennett *et al.* (2004). There are two reasons for adding this decision-making step. First, the environment-to-health link would have to be taken into account in any event in forming expert judgements on the effect of changes of environment on welfare. As already stated above, it is better to make these kinds of judgement explicit than have them implicit. The second, and equally important reason, is that introducing a set of projections of health state provides a much more accurate means for validating the functioning of the policy evaluation tool than using an aggregate measure (or measures) of welfare change. Health state, in this system, can be described using a set of indicators recognised by industry which can subsequently be directly measured on farms or in the slaughter house. For example, suppose that the policy evaluation tool projects that changing the rearing environment for poultry (for example, decreasing stocking density to a maximum of 15kg/m square) reduces incidence of injurious pecking by 20% in free range flocks; on-farm inspections can subsequently validate whether this has actually occurred. However, if only an aggregate welfare score were available, how easy would it be to validate that, say, a 5% increase in poultry welfare had occurred? The issue of validation of projected outcomes will be revisited later. Table 8 shows the environmental measures used in the WQ system, set against those employed by Bennett, *et al.* (2004). Table 10 shows the health (and behaviour) outcome measures employed by the same systems. Versions of these measures (or indicators) would be used to delineate the rearing environment-to-health link in the policy evaluation tool.

The process of linking environmental states to health (and possibly behavioural) outcomes is not easy, for the reason that singular aspects of the environment, as represented by individual measures, are known to have a multiplicity of health and behavioural impacts. For example, in poultry rearing, feeding regime is known to affect a whole range of health-related outcomes, such as growth rates in broilers, egg number and weight in layers, body and feather condition, levels of aggression, feather pecking, disease resistance etc.

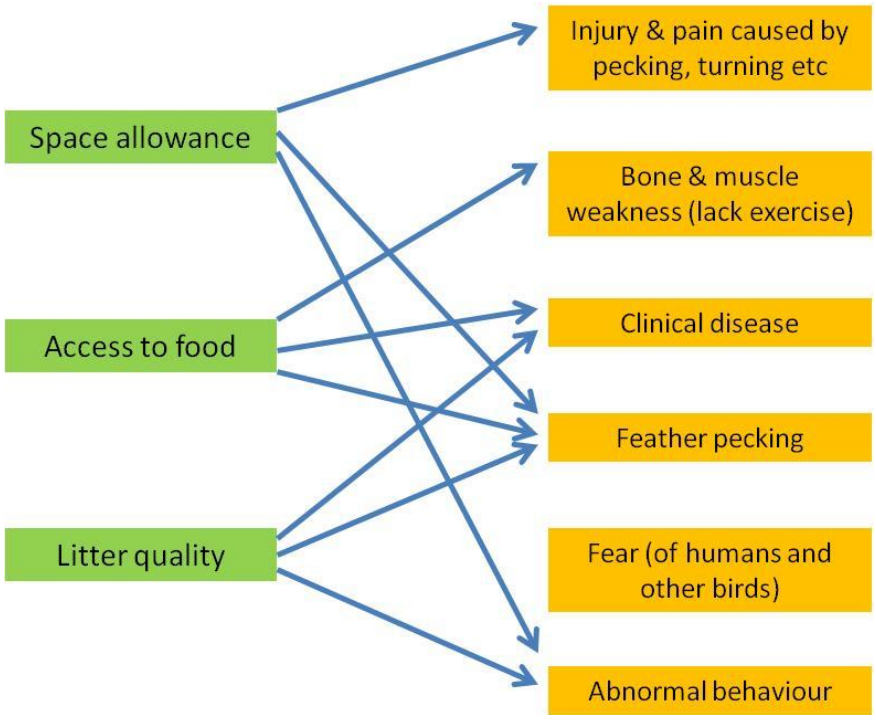
In order to make the mapping of the linkages between environment and health factors more manageable, some simplification will be required. A suggested approach to achieving this is given below, although an expert group may be able to develop a better system.

Table 8. List of environmental measures developed by the WQ project and Bennett *et al*. (2004)

List of environmental measures developed by the WQ project	List of environmental measures developed by Bennett *et al*. (2004)
Flooring	Space constraints to normal, undisturbed preening, turning, wing flapping, perching etc
Light patterns / spectral frequency / intensity	Access to litter for dust bathing, preening, and hence reduction in abnormal behaviour
Natural daylight	Access to appropriate nest sites for egg laying, and reduction in abnormal behaviour
Feeder	Access to perches for resting and sleeping
Drinkers	
Nest boxes	
Space	
Alarm systems	
Biosecurity resources	
Litter	
Perches	
Manure removal	
Placement of resource / maintenance	
Pest control	
Parasites	
Ventilation	
Predators	

Rather than try to map all known links between environmental factors and health and behavioural factors, the following approach is proposed. For each environmental factor in turn, an expert panel would be asked to identify the three most important health/behavioural linkages and these would then be mapped. Once this process has been followed for all environmental factors and all linkages had been identified, the panel would generate weights for each link to each health measure reflecting their importance. A worked example of this outcome is given in Figure 2 below.

Figure 2. An illustration of the mapping of the links between environmental factors and health factors, based on each environmental factor linking to a maximum of 3 health factors



In the example above, feather pecking (as a health outcome) is deemed to be affected by space allowance, access to food and litter quality. The panel might weight each of these drivers in terms of importance (for feather pecking) on a three-point scale as: food=3; litter quality=2; and space=1. Each environmental measure is then scored for policy change on a three point scale of 0 to 2, reflecting no change, moderate change and major change. The scores multiplied by the weights would yield a maximum score of 9 for each environmental measure, but the maximum score for each health factor would be 18, i.e. 2 X 9.

Following the example above, suppose a new policy proposal had two environmental outcomes: (i) reducing stocking density by 10% for all laying flocks; and (ii) requiring a minimum of 1 inch of friable litter to be available at all times. Based on the process outlined above, the following health scores would be recorded (Table 9).

Using the illustrative calculations shown in Table 9, feather pecking, and abnormal behaviour, would go forward into the process of measurement of welfare, both with scores of 5. However, these two health and behaviour measures would themselves be weighted, reflecting their importance in determining welfare. The process of linking health and behavioural measures to welfare is discussed below.

Table 9. Worked examples of the use of scores and weights in estimating the impact of environmental factors on health measures

Feather pecking		Weight reflecting the importance of each environmental factor as a driver of feather pecking	Score for environmental measure (reflecting policy outcomes)	Aggregate score
Environmental factors (drivers)	Space	1	1	1
	Access to food	3	0	0
	Litter quality	2	2	4
				5

Abnormal behaviour		Weight reflecting the importance of each environmental factor as a driver of abnormal behaviour	Score for environmental measure (reflecting policy outcomes)	Aggregate score
Environmental factors (drivers)	Space	3	1	3
	Litter quality	1	2	2
				5

(iii) Linking health states to welfare

It has been argued convincingly in various quarters, including the WQ project, that measures of animal health are far better indicators of welfare than are environmental measures. For this reason, the policy evaluation tool bases its welfare assessment solely on health measures. This is an adaptation to the WQ system, which currently includes some environment measures where reliable health measures do not (yet) exist to represent some issues. For example, because there is no appropriate health measure to capture thirst, except where dehydration becomes severe, the WQ system includes with its health measures an environmental measure capturing access to drinking water (this might be expressed in terms of presence, functioning and cleanliness of drinkers per given number of animals etc).

The health (and behaviour) measures employed in the policy evaluation tool would need to be species specific. Table 10 shows animal health and behaviour measures developed for the WQ project. The more generalised set of measures employed by Bennett *et al.* (2004) are presented by way of comparison. The animal health measures in Table 10 are confined to those developed for poultry for sake of brevity.

Table 10. List of health and behavioural measures developed by the WQ project and Bennett *et al*. (2004)

List of poultry health measures developed for the WQ project	List of poultry health measures developed by Bennett *et al*. (2004)
Absence of injuries (Lameness; Hock burn; foot pad dermatitis; Breast burns) Absence of pain caused by management procedures Absence of disease (On-farm mortality; Culls on farm; Ascites; Dehydration; Septicaemia; Hepatitis; Pericarditis; Abscess) Expression of social behaviours Good human animal relationship Absence of fearfulness	Injury and pain caused by being pecked, turning, wing-flapping, perching etc. Bone and muscle weakness Clinical disease Feather loss (caused by feather pecking, treading and abrasion); Injurious pecking; Fear (caused by humans and other birds) Abnormal repeated behaviour

The WQ protocol is currently being developed for more species and work is ongoing to develop a version of the protocol that can readily be rolled out for on-farm welfare evaluation (this might involve some simplification to reduce the resource and data burden associated with using it), for example for the purposes of monitoring of compliance with standards contained in accreditation schemes that have an animal welfare component. It would seem reasonable then to base the animal health measures of the proposed welfare measurement system on those developed for the WQ as far as is possible. By doing so, this would mean that data gathered, in the round, through use of the WQ protocol in the future could be used to validate the functioning of the new policy evaluation tool.

If the WQ health measures list is used in the new policy evaluation tool, another possibility that perhaps should be examined is the use of the technical coefficients derived by the WQ project to generate the final welfare score(s).

By doing this expert judgement would be needed simply to supply the data on environment and health outcomes (the first two steps out of three in the chain) while the WQ coefficients

would form the last link on the chain, i.e., the health-to-welfare link, in order to estimate a welfare score.

The main benefit of this approach would be that once expert judgements had delivered health outcomes on the basis of environmental state, no further expert judgement would be necessary in order to operate the tool, as the WQ technical coefficients would do the rest. These technical coefficients consist of a series of weights that rank the various health, behavioural and environmental measures in terms of their importance for welfare, so that they can be combined into aggregate measures (or dimensions) of welfare.

Obviously, because the policy evaluation tool would not contain environment measures at this stage, the WQ weights for environment measures would have to be stripped out and, perhaps, the weights for the health measures re-scaled.

There are two further potential problems with using the WQ methods in this way. The first is a consequence of the detailed nature of the health measures used in the system. Indeed, some of the measures shown in Table 10 are themselves aggregates of component measures. Because of this complexity, and the specificity of the individual measures, it might be difficult to generate health outcomes for individual measures with the precision that is necessary.

However, if this proved to be a problem in practice, an adaptation of the WQ measures could be employed to aggregate health measures to simplify the cognitive burden associated with projection of health outcomes. Also, there exists the possibility that, in the short-term, the existing WQ protocols will be simplified in to make them more workable, in the light of experience gained trialling them on farms.

If the WQ system of structure and weights is used to create the health-to-welfare link, then the WQ approach of deriving 16 'Criteria' values can also be deployed, and these could further be combined to yield the four 'Principle' welfare measures. This, in itself would provide the user of the Policy evaluation tool not only with general data on the magnitude of welfare change (over populations of different species), but also interpretive detail on the nature of these changes. To arrive at a single aggregate measure of welfare, the four 'Principle' dimensions (good feeding, good housing, etc), being scored on a pseudo-cardinal scale, can be mathematically combined. However, in doing so, attention should be paid to the decision rules developed for doing so in the WQ project, which reflect a group decision that dimensions of welfare should not be allowed to compensate for one another, when a score on one dimension falls below a given threshold. When constructing the policy evaluation tool, at the point that the expert group reviews the weights employed by the WQ system to derive the welfare measures, they should also extend their analysis to consider this non-compensation rule.

This process of treating environment factors as causative drivers and mapping and weighting their associations with health and behaviour outcomes, is redolent of neural network

analysis. By this approach, primary environment scores (essentially the interpretations of policy) are carried along webs of associations, attracting a variety of weights, to construct a complex picture of health and behaviour outcomes, which in turn are weighted and contribute to a measure of overall welfare. The final picture that emerges may be relatively simple to express, but is the net effect of a complex array of interacting forces, each operating according to a quantifiable (and quantified) set of assumptions.

6.6 How would the policy analysis tool work in practice?

As stated above, the policy evaluation tool would be a process as much as a product. The product would constitute the data, documentation, IP and operating structures, and possibly software files, while the process reflects the interaction of expert groups and end-users with the product for various purposes. As indicated above, there would be four types of process:

- Construction
- Application
- Validation
- Updating

Each of these process stages would involve engagement with either expert stakeholders, end-users, or both, although some of the process, for cost reasons, would only be undertaken infrequently, or perhaps just once at the outset.

Construction

This document has hopefully achieved a number of objectives, i.e. it has detailed the requirements of the policy evaluation tool; assessed the potential for basing the tool on existing welfare measurement systems (both methodologies and data); and proposed a methodology for the construction of the new tool. However, the actual construction of the new tool would have to be carried out within the context of a new research project, with the researchers actively engaging with expert stakeholders. The work of the construction project would follow a number of steps:

(i) The proposals made in this project for the design and the methodology for constructing and operating the policy evaluation tool would need to be reviewed by the expert panel and refinements made in light of feedback.

(ii) The project team would identify appropriate lists of environment (incl. management) and health and behavioural variables for use as indicators for all farmed species (based on those identified in this project) and these would be reviewed by the stakeholder group.

(iii) The primary data required for the tool would be generated by the stakeholder group using a form of Delphi exercise for a single species in the first instance. There would be two separate elements to this: first, mapping and weighting the links between environmental factors and animal health and behaviour; and second reviewing the weights provided by the WQ system for the health-to-welfare link and suggesting necessary revisions that would be necessary in light of the removal of the environment measures.

(iv) Roll the process out for the remaining species.

Application

Once the policy evaluation tool is constructed, the IP contained in the tool, together with the associated documentation and datasets would reside with Defra. However, in order to analyse new policy proposals, a second round of engagement with an expert group will be necessary. This second stakeholder consultation would have two purposes: first to interpret the provisions of new policy in terms of environment (and management) outcomes; and second, to generate scores for the environment measures. At this point the involvement of the expert group can be terminated (although they can be asked to carry out some validation tasks, see Validation section below) and the tool can be mechanically worked (based on embedded weights and processes) to generate: first, projections of changes to health and behaviour states; and then a welfare score. This process of engaging with the stakeholder group and operating the model could be carried out by Defra staff, or by using a commissioned third party. Indeed, with growing experience of the tool Defra may feel that they have much of the expertise needed to interpret the environmental outcomes of policy provisions in house.

Validation

Validation of the robustness of the policy evaluation tool and its projections can take place at three stages, first at the point of construction, second when operating the model during scenario analysis, and finally at future points in time, when counter-factual data becomes available on the outcomes of policy changes.

Once the policy evaluation tool is constructed, it would be sensible validate it, i.e. check that it is operating properly and that outcomes are reasonable. To achieve this, a number of test-case policy scenarios should be run. These scenarios might be policy measures that have already been introduced, where outcomes are already known, or novel policy measures, so designed to test a range of possible outcomes, for example, scenarios at a range of policy extremes. The expert group would help in the design of these scenarios and review outcomes and on the basis of their assessments adjustments to weights within the tool could be made.

During the operation of the tool, experts will be used to interpret policy provisions in terms of changes to environment state and the tool will then be run using this input. The outputs of the tool could then be fed back to the expert group for validation. The group could be asked, for example, to examine the aggregate changes to health states derived by the model from the putative changes to environment. If they thought that these aggregate health changes would infeasible, this would signal the need to adjust the model coefficients to constrain outcomes within more realistic bounds.

Once the tool has been operating for a time, it would be possible to examine its performance by comparison with the counter-factual (i.e. real world outcomes). Counter-factual data might be obtained from a number of sources, including specially commissioned surveys of farms, through use of expert opinion on outcomes, and through the use of third-party data, such as

from the Animal Health and Welfare Inspectorate, or data from accreditation scheme inspections that might be using the WQ protocols, etc. As has been stated above, the availability of intermediate tool outputs, i.e. projected changes to states of animal health and mental state make this type of validation exercise possible, because this type of outcome can be directly measured at the farm level. As has also been noted, the choice of health and behaviour measures for use in the tool is key to the type of data sources that can be used in future validation exercises. Basing the health measures contained in the tool on those deployed in the WQ protocols seems sensible, in view of the likelihood that these will be commonly used in the future for farm-based assessment of welfare (perhaps under accreditation schemes).

What should be stressed here, however, is that this policy evaluation tool will not replace on-farm monitoring of animal welfare. Rather, it will facilitate ex-ante assessment of policies and is designed to function within a policy design, implementation and review framework, in conjunction with existing processes, such as ex-post assessment of welfare outcomes (which will themselves provide counter-factual data that can be used to calibrate the evaluation tool).

Updating

As the policy, farming and scientific environment are constantly evolving, it will be necessary to periodically update the policy evaluation tool to ensure that it remains valid for policy analysis. There will be three drivers of updating: (i) improved scientific understanding of welfare issues; (ii) technological developments in the farming industry (resulting from new breeding technologies; new feeds, medicines and housing; and management changes driven by policy); and (iii) the presence of 'counter-factual' data on welfare derived from farm surveys (see Validation section above). Drivers (i) and (ii) are likely to operate over longer time-scales, suggesting that updating need not be very frequent. However, the most pressing case for updating will be the arrival of data from on-farm monitoring exercises on real-world outcomes of policy changes that have previously been 'projected' using the welfare measurement system. This data will provide key intelligence that the system is, or is not, functioning properly and if the latter, revision of the model to correct for malfunction should occur rapidly. At the same time, updating on the basis of (i) and (ii) can also occur. As a priority then, Defra should seek to ensure regular collection of counter-factual data in the wake of new policy implementation. By means of this cycle of validation and updating, the performance of the policy evaluation tool will improve with time and use.

The expert group

Such a tool as described is dependent on input from an expert group at various points in its creation and operation. Obviously the larger the group the more balanced and reliable will be their outputs, but the management burden will also be greater. Obviously, Defra cannot compel experts to give up their time, so to facilitate a voluntary approach the demands made on individuals should be kept to a minimum and a key element of that must be the use of experts on an ad hoc basis, rather than a requirement to commit for fixed periods, or multiple phases of work.

The key input of stakeholders has to be in the creation of the tool itself and so efforts should be made to obtain the largest and best qualified group to undertake this task. Subsequent involvements of experts, for example in the operation of the tool for policy analysis, would be less demanding and a relatively small group would suffice. The periodic validation/updating exercises would lie somewhere in between, in terms of intellectual demands and group size. Bennett et al. (2004) employed a group of 68 in a study to rate the impact on animal welfare of various welfare measures/schemes, and a group of around 50 would therefore seem appropriate for input to the construction phase.

Historically, a number of techniques have been used to facilitate the derivation of data for welfare measurement systems, such as Delphi and MCA techniques, and use should be made of these in this case. Obviously, the more rigorous the techniques, the more reliable the data that will be derived, but doing so may increase the cognitive burden on the expert group. Among these techniques, Delphi is perhaps the most valuable as it allows individuals to observe the responses of others, and revisit their own judgements in light of this. The great benefit of this is that it, in theory at least, it allows individuals with inferior knowledge on an issue to be influenced to improve their performance by reference to the judgements of others, who hopefully have more knowledge. However, what can happen in situations where there is widespread ignorance or uncertainty among the expert group is that there is regression to the mean in revision rounds, without any real improvement in accuracy. As such situations are quite likely in this case, techniques should be deployed to deal with this, such as asking experts to rate the confidence they have in their judgements, or analysis of the coefficients of variation in first round responses and downgrading the weights associated with measures that score poorly in this regard.

Ex post evaluations

The policy evaluation tool outlined above has been primarily designed to permit ex ante policy evaluation where the input data to the tool are generated either by expert opinion (on the impacts of policy changes on the rearing environment), or from trails of new husbandry practices or production systems, where such data are available. However, the tool can also be used for ex post assessments of welfare outcomes by use of health state data collected from farms that have implemented new policies. By the collection of data on a range of health indicators from farms by means of survey, these data can be input to the tool, which will generate welfare scores for farms or livestock sectors in the same way that the WQ protocol does. The tool can thus provide ex ante projections of welfare gains resulting from policy proposals and ex post estimates of actual welfare gains using the same welfare metric.

6.7 Operation of the policy evaluation tool –a case-study using partial data

Introduction

The purpose of this case-study is to show how the policy evaluation tool proposed in this study might be used for ex-ante evaluation of the impacts on animal welfare of proposed

changes to policy or regulation governing the rearing of farmed animals. It is assumed for these purposes that the tool is already constructed and ready for use in policy analysis, based closely on the methodology of the Welfare Quality Protocol. For illustration purposes this worked example uses a real-world policy development i.e. the recent EU directive on broiler production (EU Directive 2007/43/EC), but the analysis of this policy change is notional and partial. The purpose of this case-study is not to analyse the policy but simply to illustrate the use of the policy evaluation tool, and in particular the following: the role of the expert group, the nature of the expert input, the time-table for running the evaluation tool, the functioning of the evaluation tool itself and the types of outputs that can be expected.

A case study policy scenario

The policy evaluation tool will be able to evaluate the welfare impacts of any proposed policy (or regulation), or policy change, that impacts on the way in which farmed animals are reared. These policies can be specified in any detail, from fairly undeveloped ideas containing broad-brush proposals, to fully-developed and detailed draft official legislation, as well as recently implemented policy. The only requirement is that there must be sufficient detail in the proposed policies to allow their interpretation, by an expert group, in terms of changes to the physical/management environment in which animals are reared. For the purposes of this case study we will use, as a basis for further analysis, EU Directive 2007/43/EC. In England, this directive will mean:

- A reduction in the maximum stocking density from 42 kg/m^2 to 33 kg/m^2 as standard, or 39 kg/m^2 if certain additional requirements are met
- The collection, at slaughter, of data for bird physical welfare indicators, to be used to trigger alerts to animal keepers and Animal Health if values reach certain thresholds
- A requirement for keeper training in various aspects of livestock husbandry and welfare practices, incl. physiology, feeding, animal behaviour, transport, emergency care and culling, bio-security etc.

The Directive also specifies standards for the production environment, such as lighting patterns, ventilation, air quality parameters, humidity and temperature, litter quality, house noise levels, cleaning schedules, inspection intervals for animals, farm plans, alarm systems and feed withdrawal times.

For our purposes we will focus on just one element of the directive, the maximum stocking density reduction.

Source data

The policy evaluation tool requires two types of data. First, a set of policy proposals and second, the interpretation of these policy parameters in terms of their impacts on the physical and management environment in which animals are reared. This interpretation can itself come from two possible sources. First, from trials data, i.e. where policy provisions are implemented on a sample of farms to evaluate their impacts (or the tool could be applied

post-policy when data from policy implementation could be available). Second interpretation can come from the judgements of an expert group, who project the likely impacts of policy provisions on the rearing environment based on their knowledge of livestock production and/or animal welfare. As it will be more often the case that trials (or post-policy) data are not available, for the purpose of this case study the operation of the policy evaluation tool will be based on the supply of scenario data from an expert group.

The expert group

The size and composition of the expert group will in part be determined by the nature of the policies that are being examined. Broad policies impacting multiples issues in multiple species would require a larger group with broader expertise than more tightly targeted policies. Group size would likely vary therefore between 6 and 10. In terms of composition, the group will need to contain a core of welfare scientists and welfare assessors, with other individuals brought in as needed to provide particular expertise. The primary role of the expert group in this context is to interpret the proposals for policy in terms of impacts on the environment (and management). This is achieved by the generation of scores for the environment measures contained in the model reflecting the impacts of the policy provisions on these aspects of management and the physical environment. However, experts would also be asked to provide estimates of the impacts on selected welfare outcome measures for model validation purposes.

Members of the expert group might be drawn from: various government ministries, Animal Health, VLA, Veterinary professionals, academic institutions, NFU, breed societies, RSPCA, etc. The expert group would be sent a briefing document, on recruitment, detailing the provisions of the policy proposals and details of the data requirements of the policy analysis tool, i.e. a list of the environmental variables contained in the model and the baseline values of these. A group meeting might then be convened. The meeting would last at maximum half a day and would begin with a round-table discussion of the policy provisions, followed by a first round of voting on changes to the values of the environment variables. A second round-table discussion will focus on outlying scores and this will be followed by a second round of voting. The second round votes will be averaged and these averages will go forward for entry into the tool. At this point the expert group can either be dismissed, and the evaluation tool can be mechanically worked (based on embedded weights and processes) at a later date, or the toll can be run at the meeting so that the expert group can comment on the outputs. The outputs would include: (i) projections of changes to health and behaviour states; and (ii) a global welfare score. Alternatively, the judgements of the expert group could be elicited by email, or via a web-based forum. This would eliminate meeting expenses and might encourage participation.

A worked example

The policy definition stage

The case study policy (EU Directive 2007/43/EC) impacts solely on broilers and only one dimension of the policy will be examined, i.e. a reduction in maximum stocking density from 42kg/m square to 33kg/m square.

The expert group stage

Table 11 below illustrates the type of data that might be elicited from the expert group. The table contains a selection of the environmental variables that would be contained in the tool and notional baseline scores for each variable (these baseline scores would have been agreed by expert stakeholders during the construction of the tool). The environmental variables included in Table 11 are appropriate for the limited policy dimension (i.e. stocking density) being examined in this case study. Also contained in the table, in blue text, are notional values representing the outputs of the second round of voting by the expert group on the changes to environmental states that would result from the maximum stocking rate reduction.

Table 11. Sample of model environmental variables with baseline and expert group-adjusted values for policy scenario (variables ranked on a 0-10 scale, where 10-best possible condition)

			Litter quality		
	Noise levels	Space	Cleanliness	Moisture	Air quality
Baseline score	5	6	6	5	7
Scenario score (round 2)	5.2	6.5	6.5	5.3	7.1

The desk-top modelling stage

The desk-top modelling stage takes the new data for environmental variables derived from the expert group and interprets these revised values in terms of a suite of health and behavioural indicators. These calculations are entirely automated and occur within the evaluation tool. Table 12 illustrates the operations carried out on the environmental data within the evaluation tool for the case study environmental data from Table 11 above. Obviously, the final policy evaluation tool will contain much more complexity and will have multiple elements covering different species (although separate tools for each species might make handling the tool easier). The tool should be simple enough to construct in a spread-sheet, or perhaps using a systems modelling package such as STELLA.

Table 12. Spreadsheet-style representation of the calculations carried out in the policy evaluation tool – the baseline case.

BASELINE	Env. Variable Score (from expert group)	Feather condition		Skin condition		Foot pad dermatitis		Abnormal behaviour		
		wgt	Weighted score	wgt	Weighted score	wgt	Weighted score	wgt	Weighted score	
Space	5	1	5	6	30	0	0	7	35	
Noise level	6	0	0	0	0	0	0	5	30	
Litter cleanliness	6	8	48	6	36	6	36	6	36	
Litter moisture	5	8	40	5	25	8	40	4	20	
Air quality	7	0	0	3	21	0	0	1	0	
Re-based total (sum/10)			9.3		11.2		7.6		12.1	
Weight of contribution of each physical measure to overall welfare		2		6		7		4		**Global welfare score**
Weighted score for each measure		18.6		67.2		53.2		48.4		**184.6**

Note: wgt is the weighting of importance (on a score of 0-9) of each environmental variable in determining the score on each health and behavioural indicator. All weights are shaded in red.

Table 13. Spreadsheet-style representation of the calculations carried out in the policy evaluation tool – the policy scenario.

POLICY SCENARIO	Env. Variable Score (from expert group)	Feather condition		Skin condition		Foot pad dermatitis		Abnormal behaviour		
		wgt	Weighted score	wgt	Weighted score	wgt	Weighted score	wgt	Weighted score	
Space	5.2	1	5.2	6	31.2	0	0	7	36.4	
Noise level	6.5	0	0	0	0	0	0	5	32.5	
Litter cleanliness	6.5	8	52	6	39	6	39	6	39	
Litter moisture	5.3	8	42.4	5	26.5	8	42.4	4	21.2	
Air quality	7.1	0	0	3	21.3	0	0	1	0	
Re-based total (sum/10)			9.96		11.8		8.14		12.91	
Weight of contribution of each physical measure to overall welfare		2		6		7		4		**Global welfare score**
Weighted score for each measure		19.92		70.8		56.98		51.64		**199.34**

Interpretation stage

Based on these partial calculations, the improvement in overall welfare of broilers from the changes to maximum stocking density would be 8 points. Converted to a 100-point scale, i.e. percent, this would be 8%. However, the policy evaluation tool also generates data which allows the construction of a more detailed interpretive picture of these welfare changes. For example, it can be seen from comparing the data in the two tables above that the stocking density changes would lead to only a minor improvement in feather condition, while the greatest improvement would be seen in (reducing incidence of) abnormal behaviours.

Validation stage

Validation will take place in two ways. First, if the desk-based policy evaluation tool can be run using the expert group data at a meeting of the experts (and there is no reason why it cannot), it would be possible to present the results to the group to solicit their reaction. Members of the expert group might then be asked to judge whether the scale of the estimated global improvement in welfare is consistent with their expectation. Additionally, the interpretative data can be used to identify whether unexpected outcomes are occurring, so that adjustments might be made.

The most compelling validation of the performance of the evaluation tool will occur when, at some undefined point in the future after the implementation of policy, data can be collected from farms using the same health and behavioural measures employed in the tool. These data can themselves be entered into the policy evaluation tool and the resulting projection of welfare change can be compared to the change projected using the data from the expert group. If the projected changes in health and behaviour measures are also seen on farm then this validates model performance; if not, then adjustments will have to be made.

Summary of stages of operation

1. Construct policy scenarios
2. Recruit appropriate expert stakeholder group
3. Issue a briefing document to the expert group
4. Convene a meeting of the experts
5. Solicit data on likely changes to production environment resulting from policy provisions (using a Delphi -type approach)
6. Enter environmental change data into the desk-top policy evaluation tool and run it
7. Interpret the health, behavioural and welfare outcomes
8. Possibly feed outcomes back to stakeholder group
9. Make any necessary adjustment to model weights or environmental input data and run the evaluation tool again

7. Economic valuation of improvements to the welfare score of farm animals

As people's beliefs, attitudes and perceptions change so society's valuation of farm animal welfare changes. The use of stated preference methods can take account of this by informing respondents about livestock production practices and their welfare implications prior to eliciting their WTP to improve animal welfare (Bennett, 1995). This means that the values obtained from a stated preference study reflect those of the informed respondent and not the population at large who may not be so informed.

7.1 Use of welfare score in valuation study

The economic valuation method recommended for the valuation of animal welfare involves using the animal welfare score outlined previously in Section 6 based on the Welfare Quality system of farm animal welfare assessment. In theory, such a scoring system could be developed and used in relation to all types of animal and not only farm animals. It then uses choice experiment and contingent valuation methods to elicit valuations from citizens for changes in the levels of welfare scores for different farm animal species.

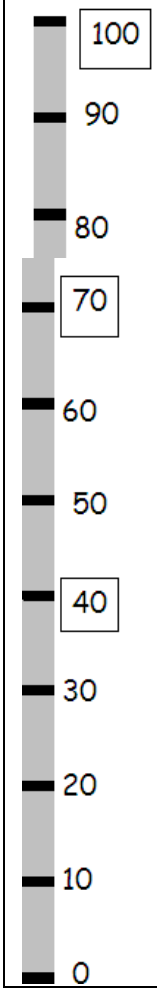
A single score/measure of animal welfare is considered necessary in order to clearly communicate to citizens the change in welfare status of the animals for which they are asked their willingness to pay. The current Welfare Quality system of measurement aggregates scores into four principle areas (good feed, good housing, good health and appropriate behaviour) which are then used to assign the welfare status of animals into one of four different classes (excellent, enhanced, acceptable or not classified). The cognitive effort for laypeople to consider animal welfare on a four-dimensional scale is substantial and is considered an unrealistic expectation in the context of willingness to pay surveys of citizens. Moreover, assignment of welfare status of animals to just one of four classes (levels) does not allow for marginal changes in welfare that might result from a policy or other measure that impacts on welfare. In contrast, a single score is probably the simplest way to communicate the welfare status of animals on a continuous scale (e.g. from 0 to 100), if it is communicated as a credible, holistic measure.

An exploratory research study applying this approach is reported below. Despite the relative technical complexity of the production of a welfare score, with various weightings and other considerations explained in Section 6, the score itself needs to be simply and clearly explained and presented to respondents when using stated preference methods to elicit people's willingness to pay for an increase in welfare of specific farm animal species, shown by an increase in the welfare score. Box 1 below shows how this information was communicated to respondents in the study. Focus groups with citizens found that the welfare scoring system described in Box 1 was both comprehensible and credible to them.

Box 1

Welfare measurement and welfare score

In the last year in the UK, 2.6 million cattle & calves, 9.5 million pigs and 798.3 million chickens were killed for meat production. Animal welfare scientists and veterinarians now have developed a system for measuring the welfare of individual animals that takes account of the varying needs of different species, ages etc. The system scores the extent to which the needs and wants of the animal are met and results in an overall welfare score on a scale of 0-100 which accurately represents the welfare of the animal in terms of its freedom from hunger, thirst, discomfort, pain, injury, disease, fear and distress, and the extent to which the animal can express natural behaviours and has a happy and contented life. A score of zero would denote extreme suffering whereas a score of 100 would denote the highest level of welfare that could possibly be achieved. The system applies over the entire life of the animal from birth to slaughter and involves regular independent monitoring of the animal's welfare throughout its life.



100

The best welfare possible

Animals are subjected to no harm and experience no suffering e.g. from disease and injuries. All of their wants such a feed, water and expression of natural social behaviour are met.

70

A good life

Animals are subjected to little harm and experience only minimal suffering e.g. from disease and injuries. Most of their wants such a feed, water and expression of natural social behaviour are met.

40

Legal minimum

This is considered to be the minimum acceptable level of welfare and the level to which most UK farmers are currently working.

Welfare levels below 40 are generally not acceptable in the UK. However, there maybe other countries where animals are produced to scores below 40.

The scoring system presented to respondents in the survey valuation exercise can be related to the Welfare Quality system as summarised in Table 14 below. Welfare Class 0 (i.e. 'unclassified') is represented by a welfare score of less than 40, where 40 is defined as representing the level achieved by compliance with legal minimum standards for welfare. The score bands representing the remaining three classes are somewhat arbitrary. A score of 40 or more would represent Welfare Class 1 or above. For example, Welfare Class 1 might be 40-60. Welfare Class 2 61-80 and Welfare Class 3 81-100.

Table 14. Relationship between the Welfare Quality welfare classes and the welfare scoring system

Welfare Quality® Index	**Welfare score**
Welfare class 3	81-100
Welfare class 2	61-80
Welfare class 1	40-60
Welfare class 0	0-39

In terms of the Farm Animal Welfare Council's (FAWC, 2010) recommendations concerning 'a life worth living' and a 'good life' for animals, a score of less than 40 might represent a life not worth living and one of 40 or above a life worth living. It is debatable at what score animals would be considered to have a good life, and clearly with a welfare scale of 0-100 there are different degrees of good life that are possible.

It is possible, that the welfare attributes of livestock production practices (and changes to them) could be reflected by the animal welfare score and appropriate changes to it. This could be done either using welfare assessment research findings based on detailed scientific welfare assessments of animals on farms, during transport, in markets and at slaughter and/or by means of veterinary expert assessments using the Welfare Quality protocols or their variants. However, further research is required to explore the feasibility of this approach.

7.2 CE and CV willingness to pay surveys – methods

The welfare scoring system was then used as the basis for CE and CV surveys to elicit values of improvements in animal welfare signified by an increase in welfare score. CE and CV questionnaires were designed. Both comprised the following structural elements: questions about attitudes, opinions and uses; information about the welfare score; two valuation scenarios; and debriefing and socio-economic questions.

The questionnaires were pre-tested both in a focus group and in a number of personal interviews and then piloted on a sample of 50 respondents separately for each questionnaire. Open-ended willingness to pay levels were also elicited from a sample of shoppers to help define the WTP bid amounts used in the questionnaires. Two separate surveys were then undertaken. A random sample of citizens in Great Britain was used, stratified according to socio-economic group. These were then telephoned to recruit them to the survey. Those who agreed to take part were then either interviewed then if they had access to the internet or an interview was arranged for a few days later. Information about the welfare scoring system and the CE (including choice sets) or CV exercise were made available on the internet for those with access or sent by post prior to interview. Around 300 respondents participated in both surveys and had completed questionnaires. For each survey, over 2,700 people were contacted, of which around a quarter agreed to be interviewed (some of which were not then used once the 300 quota had been reached for each survey).

In both surveys respondents were asked their WTP for (i) meat products from animals with a range of welfare scores in terms of the monthly and yearly additional cost of doing so (i.e. due to higher prices) and (ii) legislation which ensures that all farm animals of a particular species in the UK have a specified welfare score as a minimum.

7.2.1 CV survey

WTP for higher welfare meat products

Respondents were told 'Assume that in your usual food store there are meat and meat products with high welfare scores. If you buy this meat with a welfare score above the legal minimum 40 your monthly food bill will rise'. They were then asked 'If the meat had a welfare score 60, would you buy it for the extra cost of….?' Three different amounts were presented to respondents and they had to choose whether they would Definitely buy, Probably buy, Definitely not buy, Probably not buy the meat or if they Didn't know. This question was then repeated three times more with different welfare scores and different WTP amounts.

WTP for welfare legislation

Respondents were told 'Assume that the Government introduces legislation to ensure that all farm animals in the UK have a certain welfare score. The cost of such a scheme would be offset by a tax that has to be paid by all householders including you. The revenue from this tax would only be used to pay for this welfare legislation. In this case, the price of meat does not increase'. They were then asked 'If legislation would ensure that all farm animals have a welfare score of at least 60, would you be willing to pay a tax of….?' Three different amounts were presented to respondents and they had to choose whether they would Definitely, Probably, Definitely not, Probably not be willing to pay the amount or if they Didn't know. This question was then repeated three times more with different welfare scores and different WTP amounts.

WTP questions included WTP for welfare scores levels of 60, 70, 80 and 90 and associated bid values respectively of £5, £11, £22; £6, £12, £24; £8, £16, £32; and £9, £18, £36 for additional monthly meat expenditure and £40, £70, £140; £60, £120, £240; £80, £160, £320; and £90, £180, £360 for the increase in annual tax.

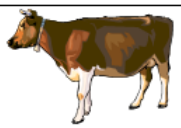
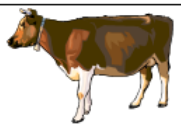
7.2.2 CE survey

WTP for higher welfare meat products

Respondents were told 'Assume that in your usual food store there is a section that sells meat and meat products with high welfare scores. The farmers who supply this meat will be monitored by the RSPCA. If you buy meat with a welfare score above the legal minimum (40) your weekly expenditure will rise. We now ask you to make six choices'. Respondents were then given some additional guidance before being presented with six different choice sets. This guidance included a reminder to them that their budget is limited and that more money spent on meat may mean less money to spend on other things. An example choice set is shown in Figure 3 below.

Figure 3. Example choice set (meat expenditure)

Suppose you could only choose one from the three baskets below. Which basket would you choose? Please circle only one!

		Welfare scores of your **meat**		
Basket A	Increase in monthly meat expenditure **£ 16.00** (£ 192 per year)	 90	 60	 70
Basket B	Increase in monthly meat expenditure **£ 18.00** (£ 216 per year)	 90	 70	 50
Basket C	Increase in monthly meat expenditure **£ 0.00** (£ 0.00 per year)	 40	 40	 40

or

Don't know.

Respondents were asked to choose one alternative only out of A, B or C (status quo) or could choose a Don't know option.

WTP for welfare legislation

The above CE exercise was then repeated but this time respondents were told 'Assume that the Government introduces legislation to improve the welfare of all farm animals in the UK to a certain welfare score. All animals would have to be kept at this welfare score. Farmers' compliance would be monitored by the Department for Environment, Food and Rural Affairs and the RSPCA. These monitoring efforts as well as farmers' investments in welfare-improving measures have a cost. This cost would be offset by a tax that has to be paid by all households including yours. The revenue from this tax would only be used to pay for the implementation of this welfare legislation. Again, please make six choices'. Again, some additional guidance was given to respondents prior to them making their choices including reminding them of their budget limitations.

The choice sets were similar to that shown above except in the left hand column an additional annual tax amount was given instead of an increase in meat expenditure (except for the status quo scenario C where no additional tax was required).

The choice sets contained 4 attributes: welfare score for beef cattle, welfare score for chickens and welfare score for pigs - with score levels of 40, 50, 60, 70 and 90 used - and a price attribute which was either the additional price of meat (at £0, £6, £8, £12, £16 and £24 per month extra) or additional annual tax for the legislation (at £0, £70, £110, £150, £190 and £240).

7.3 Main results

Data from both surveys were coded and entered onto a spreadsheet and then analysed using appropriate statistical analysis software.

7.3.1 CE survey

In terms of meat purchasing behaviour, average weekly meat expenditure was £17.65 with 86% of respondents eating chicken, 68% beef, 54% pork and 55% other meat. Less than 5% said they did not consume meat. Thirty-eight per cent of respondents felt well informed about the way in which farm animals are treated and over 72% were concerned about the way farm animals are treated.

In terms of attitudes and beliefs, 96% of respondents agreed that we have a moral obligation to safeguard the welfare of animals. Eighty-one per cent thought that meat from animals with high welfare has better food safety, 78% that it was healthier, 71% that it had better nutritional value and 69% that it tasted better.

WTP was derived from the CE data using conditional logit and mixed logit models with a Bayesian inference estimation procedure. Tables 15 and 16 show results of the preferred

models (several models were explored) with respect to price and tax payments. All coefficients have the expected signs, positive on the species attributes showing that people prefer higher welfare levels (scores) to lower ones and negative on the price and tax coefficients showing that people prefer to pay lower prices and lower taxes.

Table 15. Transformed utility coefficients of the preferred model – price

Attribute	Mean	Std. dev.	Median
Beef	0.377	0.506	0.214
Pork	0.265	0.359	0.116
Chicken	0.313	0.400	0.224
Price	-0.755	2.159	0.000

Table 16. Transformed utility coefficients of the preferred model – taxation

Attribute	Mean	Std. dev.	Median
Cow welfare	0.429	0.536	0.271
Pig welfare	0.315	0.434	0.114
Chicken welfare	0.292	0.417	0.103
Tax	-1.638	2.590	0.000

WTP for a one point increase in welfare can be computed by scaling (i.e. dividing) each welfare coefficient by the price or tax coefficient in each model. These WTP estimates are shown in Table 17.

Table 17. Mean WTP for a one point increase in welfare (£)

	CE price monthly	CE price annually	CE tax annually
Beef/cow welfare	0.50	6.00	2.62
Pork/pig welfare	0.35	4.20	1.92
Chicken/chicken welfare	0.42	5.04	1.78

It can be seen from Table 17 that WTP is highest for cow welfare followed by chicken welfare where meat price is the payment vehicle but pigs where tax is the payment vehicle. This difference between payment vehicles is probably due to the relatively large number of consumers that eat chicken compared to pork. Tax payments are generally lower than price

payments, probably reflecting people's dislike of paying tax. Indeed, 40% of respondents explicitly stated that they objected to paying more tax whilst at the same time a low percentage (24%) trusted the government to use the tax payments solely for farm animal welfare whilst only 58% thought that the majority of farmers would comply with welfare legislation.

7.3.2 CV survey

In terms of meat purchasing behaviour, average weekly meat expenditure was £17.02 per household with 84% of respondents eating chicken, 67% beef, 57% pork and 48% other meat. Seven per cent said they did not consume meat. Thirty-nine per cent of respondents felt well informed about the way in which farm animals are treated and 76% were concerned about the way farm animals are treated.

In terms of attitudes and beliefs, 95% of respondents agreed that we have a moral obligation to safeguard the welfare of animals. Seventy-five per cent thought that meat from animals with high welfare has better food safety, 78% that it was healthier, 72% that it had better nutritional value and 72% that it tasted better. Whilst this shows there was considerable overlap in the concepts of food safety, healthiness and nutritional value, in this context it does not matter. However, it would if we were to try to value these as separate attributes.

These findings are very similar to those from the CE survey.

WTP was derived from the CV data using an ordered probit model with a Bayesian inference estimation procedure. Tables 18 and 19 show results of the preferred models (several models were explored).

Table 18. Mean CV WTP estimates for high welfare status meat

Score	Monthly	(std. dev.)	annual	n
60	19.31	(1.676)	231.72	142
70	26.98	(1.543)	323.76	136
80	37.12	(1.840)	454.44	142
90	31.88	(1.645)	382.56	136

Table 19. Mean WTP estimates for welfare legislation

Score	Annual	(std. dev.)	n
60	163.71	(11.535)	150
70	204.02	(11.204)	150
80	297.55	(17.676)	150
90	260.33	(17.717)	150

Table 18 shows that respondents had a WTP of over £19 per month (nearly £232 per year) for meat with a minimum welfare score of 60 (i.e. compared with the current legal minimum of 40) and progressively higher amounts for meat of welfare scores 70 and 80. However, WTP for welfare score 90 is lower than that for score 80. Exploration of why respondents might give a lower value to a higher welfare score uncovered a number of responses to the open-ended question in the questionnaire which asked respondents to explain their responses to the WTP questions. A number of respondents made statements to the effect that a score of 90 represented a very high level of welfare for animals and that most people did not have such a high welfare level.

Table 19 shows a similar story in this regard. However, generally, annual WTP amounts are lower for higher welfare scores achieved through taxation and legislation than through prices and food stores. This is consistent with the CE findings with similar negative attitudes to paying tax and trust in the government to spend the money raised only on welfare. Some 39% of respondents objected to paying more tax whilst only 28% trusted the government to use the money raised only for farm animal welfare.

7.3.3 Comparison between CE and CV estimates

A comparison of WTP estimates between the CE and CV surveys shows similar orders of magnitude. For example, over the 40-80 welfare score range, average WTP from the CV survey was £113.61 per annum per 10 point increase in welfare score for the price payment vehicle and £152.40 from the CE survey for the three meats. WTP estimates for the tax payment vehicle were £63.20 per annum per 10 point increase from the CE survey and £74.38 per annum per 10 point increase from the CV survey.

It is worth exploring why CE and CV estimates might differ. The CE price estimate quoted above includes the separate estimates for each of the three meats (beef, pork and chicken). However, many respondents did not consume all three meats and this may mean that WTP could be lower than that quoted. The CV tax estimate may have been higher than the CE one because the former scenario asked for WTP for an improvement in welfare score for all farm animals and not just those used for beef, chicken and pork.

Reasons for the differences between price and tax payment vehicles have already been discussed.

In relation to the amount that respondents and consumers generally spend on meat purchases, their WTP for higher welfare scores appears reasonable. Average meat expenditure of respondents (both surveys) was £17.34 per week or some £901 per annum so that paying an additional 12.5% (CV estimate) or 16.8% (CE estimate) for a 10-point improvement in welfare score does not seem unreasonable.

In relation to the amount that respondents paid as tax, the average amount of direct tax (income tax, council tax and employees' national insurance contributions) paid per household in the UK was £7,178 in 2008/09 (Barnard, 2010) with £4,741 paid as indirect tax (such as VAT). Thus average annual payments for a 10-point improvement in welfare score of £63-£74 (i.e. around 1% increase in direct tax) seem entirely credible in this context.

8. Discussion and conclusions

8.1 Critique of the welfare measurement component of the policy evaluation tool

8.1.1 Welfare assessment

The policy evaluation tool outlined above has been primarily designed to permitex ante policy evaluation where the input data to the tool are generated either by expert opinion (on the impacts of policy changes on the rearing environment), or from trails of new husbandry practices or production systems, where such data are available. However, ex post assessments of welfare outcomes are also possible using the tool by use of health state data collected from farms that have implemented new policies. By the collection of dta on a range of health indicators from farms by means of survey, and the input of these data to the tool, the tool replicates the action of the WQ protocol in generating a welfare score.

The adaptation of the WQ approach to forge strong links with the rearing environment is important, because animal welfare policies are designed to control those aspects of the animal production system that can be controlled, i.e. the production environment and management. Health outcomes cannot be directly controlled, even if health outcome objectives are set within policies. Because the provisions of policies will be stated in terms of environment and management prescriptions, the policy evaluation tool must start at this point and then move through the whole system, to generate an aggregate welfare measure.

With respect of the reliance of the welfare measurement system on expert judgement for ex ante policy evaluation, there really is little alternative to this where field data are lacking and it is telling that even the multi-million Euro, multi-country, four year Welfare Quality project resorts to the use of this type of data. As has already been stated, the literature relevant to this issue is very large and reports on an enormous diversity of studies, each with different scale and scope, different primary objectives, relating to different species, focussing on different risk factors, using different metrics etc. There is no systematic framework which allows the interpretation of disparate (in all senses of the word) data from animal welfare science (and other disciplines) into the bigger picture. Without this framework, we are reliant on the interpretive understanding of practitioners who, as a body, can provide some coherence to the data. In a sense then, the policy evaluation tool becomes itself an interpretive framework, albeit that, to some extent, it can be linked directly to the empirical evidence, it is still to a large extent indirectly linked through the interpretive agency of experts.

The welfare score is not a true cardinal measure of welfare, although respondents may interpret it in this way, but rather an ordinal measure. Thus, a welfare score of 80 is certainly a very large increase in welfare from 40 but does not necessarily mean that an animal's welfare is exactly twice that when the score is 40 (for example, if a person's income doubled and they had half of the health problems that they had previously their welfare/happiness

would, *ceteris paribus*, be greater but not necessarily twice what it was before). For practical purposes this may not be an issue (since it may be a distinction that could not be easily explained to or understood by respondents) but nonetheless needs to be borne in mind in terms of the underlying theory of the method..

8.1.2 Limitations to uses

The policy evaluation tool has been developed to meet the specific requirements of the sponsor.. Although the tool is based on an existing system (i.e. Welfare Quality), the required adaptations make it bespoke, with few generalised applications beyond those stated above.

In terms of efficacy, the tool is limited by the data available and the quality of the judgments that the expert group make in its construction and operation. Various steps must therefore be taken to maximise quality, both in terms of choice of individual experts, the methodological framework in which they will make their judgements, and, perhaps most importantly, the use of real-world observation both to drive and validate the outputs of the tool (and provide calibration coefficients). In this regard, results of welfare assessments based on the Welfare Quality protocol will be particularly useful, both to inform expert judgement and as data to directly drive the tool (replacing the need for expert judgement to some, possibly considerable, extent). Without evidence derived from such assessments, experts may be limited in the extent to which their judgements are well informed and, indeed, they may be unwilling to provide such judgements under some circumstances.

8.1.3 Complexity – simplicity

As with any modelling approach, or expert system, there is a trade-off between resource availability and complexity. More complex representations of systems generally produce more accurate projections, under a greater diversity of scenarios, but with increasing complexity comes greater resource requirements, reduced transparency, a higher data burden etc. Livestock production systems, examined from the point of view of welfare are complex. Therefore, to represent the essential elements of such systems requires a minimum level of complexity, i.e. such systems exhibit irreducible complexity. The policy evaluation tool represents a balance between the level of complexity required to produce a functioning model and input requirements (this includes both financial resources and limitations to what stakeholders can be asked to contribute, both in terms of cognitive load and time commitment). However, as this tool, understood in its broadest sense, is designed to evolve, there would be opportunities to increase the sophistication of the representation of the environment-species interactions at a later date. For example, the current recommendation is that just three links be identified between each environment factor and health considerations. In reality, there may be many more links than this. More of these links could be incorporated in the future, and by building off what is already present, the cognitive requirements to do this would be marginal.

Another area where the tool as proposed here might be improved is through the introduction of internal calibration mechanisms. As presented experts would be asked to attach weights and scores to individual environmental measures in isolation and these would then be summed to provide changes to individual health measures. A useful refinement would require experts also to provide a judgement on the maximum level of change in health measures that they would expect to see resulting from all the relevant environmental changes lower down the hierarchy. This feasibility range would immediately highlight outlying health values and focus attention on the need to refine weights and scores.

8.1.4 Un-measureable management quality

Two farms, producing the same species, with identical production systems and environment conditions, can have different welfare outcomes, due to differences in management endowment. How does the policy evaluation tool deal with this? The short answer is that it doesn't explicitly deal with it, because it cannot unless appropriate field data is available.

Management quality is an unknown and unpredictable factor that cannot be reflected in the system explicitly. However it can be accounted for implicitly as an outcome. When experts are asked to make judgements about the impact of a particular environmental factor on health state, there are a range of values that any one expert could apply, depending on whether they assume perfect or less than perfect management in delivery. What the experts can be explicitly asked to do is make a judgement which accounts for the central tendency in management quality, i.e. supply a form of mean value that captures the effect of a distribution of management endowments on outcomes. One issue that this does not address, of course, is the likelihood that producers will flout the changes embodied in policy. For various reasons farmers do not always comply with welfare regulations, either because they are unaware of them, misunderstand them, require time to adjust to them, don't have the capital needed to implement them, or just wilfully ignore them. Whatever the reason, the numbers that are non-compliant can sometimes be quite high, especially in the first years after the policy changes are implemented. The policy evaluation tool cannot account for this, as the likelihood of non-compliance with policy cannot be known in advance. The tool should be understood then as projecting outcomes assuming full compliance, with average levels of associated management ability.

8.1.5 Removing the need for repeated bespoke studies to assess every change in policy

One of the stated objectives of this policy evaluation tool is the removal of the need to commission new welfare studies in order to understand the welfare impacts for every new livestock policy that is proposed or implemented. This suggests perhaps an aspiration that this project will deliver a stand-alone tool that can be taken off the shelf by policy makers, run on a desk-top to analyse a policy, and put back on the shelf again without reference to any other tools, databases, or stakeholders. If there was such an aspiration it is unobtainable. As has been stated already, the introduction of new policies introduces unknowns into the welfare equation that will always require some expert cognitive input to interpret. Cognitive

effort goes into the design of welfare policies and the use of this tool would form part of that effort. Moreover, expert judgement should be informed by scientific evidence and data. Thus, policy proposals to improve welfare may still require research or trial systems accompanied by on-farm welfare assessment (for example, based on the Welfare Quality system) or require welfare assessments on a sample of farms if the proposed policy is already practised by some producers.

8.1.6 The number of animals affected

There are two aspects to this issue: (i) the number of animals that are impacted by policy changes; and (ii) the number of animals falling into particular categories of welfare state (or changes to state).

(i) There may be situations where policy follows on from industry practice. For example where codes of good practice (or accreditation schemes) are in place and are followed by some producers but not others, and policy is introduced to enforce these practices on all. Another scenario is that some producers don't comply with policy (as discussed above). How does the tool cope with this? We have already discussed the issue of non-compliance. On the issue of policies not affecting some proportion of the population (because they already comply with policy provisions), the system deals with that in the same way that it deals with variation in management ability, by providing an average response that accounts for pre-existing conditions. In this case, expert judgements would take into account that some producers already comply with policy provision and generate a set of coefficients based on the assumption that following policy implementation, all do.

(ii) The tool as proposed will generate, for each scenario, an estimate of average welfare state, or change to welfare state. The tool cannot supply data on the welfare demographic. This may be viewed by some as a weakness, in that the numbers of animals in the extreme categories (particularly the low welfare categories) is not directly estimated. Methods by which this type of data might be inferred, post hoc, can be imagined. For example, if the shape of the distribution of welfare classes prior to the intervention is known, and it can be assumed that the shape of the distribution remains unchanged post-intervention, i.e. standard deviation scores remain unchanged, then these standard deviation scores can be used to calculate the area (and thus number) under the tail of the curve associated with low welfare scores. The problem with this approach is that constancy of the distribution pre- and post-intervention cannot always be assumed, i.e. some interventions might only affect one tail of the distribution (e.g. low welfare cases), thus changing the mean and the shape of the distribution. The only way to deal with this limitation properly would be to construct separate tools not just for species, but for welfare classes within species, with each having a distinct set of

coefficients reflecting the response of health and behavioural measures to changes in environment state.

8.1.7 The duration of improvements to welfare state

Existing systems of welfare measurement do not account for the time spent by animals under different welfare states and the policy evaluation tool does not attempt to do this either. Because separate welfare measurement tools would be constructed for different species, the problem of the different life-spans of species is in some respects circumvented. However, this problem would remain unresolved if it is necessary, perhaps in societal valuation experiments, to compare welfare states, or changes in state, between species. It does not help in this case that the same metric is used for different species, because the fact remains that a 10% change in welfare of dairy cows may apply to the animal for, say, 3 years, while a 10% change in welfare of a free range hen, will only affect them for 18 months at most. The issue of duration of impacts in cross-species comparisons would not be easy to resolve. This would require that all species are measured with reference to all other species using a metric that weights according to lifespan. However, in creating these weights who is to say that the lifespan of a laying hen is worth less to society than the lifespan of a dairy cow? This issue raises ethical questions that scientists alone (and this project) cannot solve.

What is more tractable to the scientist (albeit social scientist), is the issue of differences in the duration of welfare gains within species. For example, an intervention that yields a reduction in fear in lairage at the slaughter house may have an effect for only a few hours, while another intervention might reduce fear over the whole life-span of the animal. These kinds of distinctions can be taken into account in the expert judgements that interpret changes in the states of the health and behaviour measures in terms of welfare score. This would require a second set of weights applied to the health and behaviour state measures, one for severity and one for duration.

8.1.8 The use of a single welfare score

The tool as proposed would generate a single welfare score for each species (for each policy scenario). It does this for two reasons. First, because the project specification asks for it, and second, because the societal valuation methodology largely requires it. As will be discussed in detail in the next chapter, it would be problematic to try to generate meaningful societal valuations of welfare without the use of a single welfare score, due to the extreme difficulty of conveying complex information to the public about changes to a number of different dimensions of animal welfare.

However, the creation of a single measure of welfare raises some problematic issues that have been discussed at length in the output of the Welfare Quality project. To summarise these arguments, although the WQ system does generate a single welfare outcome, this is not based on a cardinal scale. However, the WQ approach does produce pseudo-cardinal scales for

four areas, or principles, of welfare. The authors of the WQ system argue that these principles of welfare are important in and of themselves and that they are non-substitutable, for example, bad feeding cannot be compensated for by good housing. Combining these principles, they argue, causes two problems, first it implies that substitution is possible (i.e. acceptable); and second it obscures negative extreme outcomes by averaging up, or down. We leave it to others to decide whether this problem is intractable in the case of the policy evaluation tool, but would point out that also available to the end-user of this tool is an array of health and behavioural measures that provide a detailed picture of the nature of welfare conditions and which could be combined in a number of different ways to provide aggregate measures.

8.2 Critique of the valuation framework

A critique of stated preference methods of economic valuation has been provided in the review of literature in Section 5. There are a number of critiques of the contingent valuation and choice experiment methods and their application, for example that by the Nobel Prize winner Peter Diamond, who with Jerry Hausman (1994) wrote a paper highly critical of CV (Contingent Valuation: Is some number better than no number?). Venkatachalam (2004) provides a very readable critical review of the CV method in relation to environmental valuation. Louviere *et al.* (2000) give a very comprehensive description and critique of stated preference methods, especially choice experiments. In the context of health economics, the literature displays a cautious approach to the use of choice experiment methods and recommends greater circumspection (e.g. Bryan and Dolan, 2004) but much criticism relates to the way in which certain studies apply the technique rather than the technique itself.

The major criticism of stated preference approaches to economic valuation is that they do not measure what they purport to measure. In particular, it has been argued that they do not measure people's preferences because people generally do not have well-defined underlying preferences for non-market goods, particularly those that they know little or nothing about prior to the survey. Moreover, it is argued that with the usual time frame of questionnaire surveys people do not have sufficient time to construct such preferences (the link between the 'commodity' and utility is difficult or impossible for them to evaluate) but instead will respond in some 'casual cost-benefit' manner (Diamond and Hausman, 1994) and/or will use various simple decision rules to answer the (WTP) questions posed to them.

With respect to the framework outlined here, people may not be used to thinking about animal welfare issues let alone expressing their preferences in relation to different levels of animal welfare for different species. However, they will be well used to food purchasing decisions based on the (multiple) characteristics of those foods and so in this context the choice experiment eliciting WTP for welfare in relation to a basket of meat or other animal products may more easily lend itself to people being able to express their true preferences. In contrast, people in the UK are not used to being asked about their WTP for things in terms of an increase in taxation or allocation of taxes to different needs. Another obvious factor that

may be influencing decisions using these two payment vehicles is the desire to retain choice. Under the higher taxation scenario, consumers understand that once the tax levy is applied they would be compelled to pay the additional sum. However, under the product price scenario they would retain the freedom to avoid paying the higher prices (perhaps by ceasing to purchase the higher welfare value product) if they so wished. This suggests that the former 'payment vehicle' might provide more credible responses. However, values based on willingness to pay for market goods alone may fail to adequately account for the 'public good' aspects of animal welfare (i.e. the benefits that accrue to third party others in society from knowing that someone else's consumption is promoting better animal welfare).

Criticisms of the application of stated preference approaches largely revolve around questions of validity and reliability.

In simple terms, validity refers mainly to accuracy or precision whilst reliability refers to consistency or reproducibility. Validity can be split into three types: content, criterion and construct validity. Content validity refers to the ability of the instrument used to measure the value appropriately. In the valuation framework presented above, the authors believe that there is reasonable content validity at least in terms of the price payment vehicle. This is because the scenarios and questions presented to respondents seemed to make sense to them and they appeared to be able to express their preferences without confusion or undue cognitive effort. Criterion validity requires reference to another WTP measure of the same or similar commodity. In this context, the WTP results reported above seem reasonable in the context of the prices paid by people for commodities they perceive to have animal welfare attributes, such as 'free-range' and organic livestock products (eggs, meat etc.). Construct validity has two dimensions: theoretical validity and convergent validity. There is no evidence to suggest that theoretical validity of the above framework, and the results obtained from applying it, does not apply. Responses appear consistent with economic theory – for example in terms of correct expected signs on model coefficients, explanations of responses given by respondents and diminishing marginal utility and WTP of increasing levels of the commodity (animal welfare score). Convergent validity refers to comparison of two different measures based on the same theoretical construct. In this regard, the WTP estimates derived from using the two different measures, CV and CE, are comparable and correspond closely with one another. Moreover, different estimation models of WTP produced very similar estimates.

Reliability would require that repeated measurements of WTP should produce the same value (depending on accuracy of the method) or that if people's preferences change the measure should reflect this change accordingly. The results produced from the study outlined above produced similar estimates from two different surveys using two different methods linked by a common theoretical underpinning. In addition, the WTP estimates are broadly comparable with those from a number of other CV and CE studies of animal welfare (see Section 5). Further testing of reliability could be undertaken by replication of the application of the valuation framework.

Stated preference studies are subject to a number of apparent biases that can influence estimates. These include amongst others:

- embedding (e.g. where values for a commodity appear relatively unresponsive to the scale of provision of the commodity);
- 'warm glow' (where people give a WTP because it makes them feel good);
- question order bias (where responses might systematically differ depending on the order that WTP questions are asked);
- information bias (where responses might be sensitive to the information provided either in terms of content or form);
- elicitation bias (due to the way in which WTP is elicited – e.g. single or doubled bounded dichotomous choice, the number of attributes in choice sets, use of Don't Know option etc);
- hypothetical bias (due to the hypothetical/'unreal' nature of the scenario presented to respondents); and
- strategic bias (e.g. respondents stating a high WTP that does not reflect their true WTP so that an issue is taken more seriously by policy makers).

The study outlined above did not expressly test for each of these biases but rather attempted to minimize them by careful design of the survey instruments, taking account of best practice.

One form of potential information bias is that of 'anchoring' (Tversky and Kahneman, 1974), which is a type of cognitive bias whereby individuals rely on specific pieces of information (often as initial values or starting points) to govern their thought processes or base their subsequent decision(s). The information provided to respondents when using the tool to elicit their valuations for changes in welfare scores is open to such bias. This is because the welfare score was presented as a scale of 0 to 100 with reference points given in terms of welfare descriptions or standards. Perhaps the most influential of these (although the researchers did not test for bias in this respect) was the welfare score of 40 described as being equivalent to current legal minimum standards. Respondents are likely to have used this as a base value in their consideration of other scores. Indeed, the value was presented in this way to respondents in the surveys. It is likely that respondents may have had a different WTP for an increase in welfare score from a different value (e.g. an increase in score from a different base such as 50 to 60 rather than 40 to 60). This is because (a) with a base of 40 rather than, say, 50 there is the clear inference that the current welfare of the animals is at a lower level than if a higher base was presented and (b) a move from 40-60 is a greater increase than from a higher

base such as 50-60. It is important therefore that such reference points are appropriate and are grounded in reality. For example, welfare scientists and others might be asked to agree the score that best represents current legal standards.

8.3 Conclusions

The Welfare Quality framework for measuring the welfare of farm animals is not entirely suitable for use in policy evaluation, especially desk-based ex ante evaluations. However, with appropriate modification, the Welfare Quality framework can be used to assess the welfare implications of policy changes expressed on a welfare metric, which can then be used to obtain societal values associated with the welfare benefits accruing from policy changes. In order for this to happen, it is considered necessary that the welfare status of farm animals be communicated to respondents in valuation surveys in the form of a single score on a continuous scale. In this way, marginal valuation of changes in the level of the score (which represents changes in the level of welfare) can be obtained for different farmed species. Various economic valuation techniques can be applied to estimate people's willingness to pay for improvements to welfare/welfare scores, such as contingent valuation and choice experiment methods. Clearly, the results of applying these techniques within a framework such as that presented above, need to be treated with caution and some circumspection. However, they can provide reasonably valid and reliable estimates of WTP providing the survey instruments used are carefully designed and tested and providing that results are demonstrated as repeatable and consistent with one another, as well as with available revealed preference (i.e. market) data.

The review of literature has revealed that there is no body of existing valuation data available that could be used to value any welfare changes projected by the policy evaluation tool. For this reason, it would be necessary to elicit these valuations, perhaps building upon those elicited in the recent study quoted in Section 7.

A further national WTP elicitation study, which could follow the same format as the study referenced here is needed to: (i) replicate and validate the data obtained in the existing study (i.e. for dairy, pigs and poultry); and (ii) expand the data coverage to other farmed species. Periodic replication of this type of study would reveal any trends within society leading to altered valuations.

8.4 Recommendations

The following actions are recommended.

1. The societal valuations of welfare scores can be used with a range of welfare measurement systems as long as these systems generate a single welfare measure on a cardinal or ordinal, scale. Therefore, it is timely to undertake a large-scale national willingness to pay study to derive societal valuations of changes to welfare scores for a range

of farmed species. The outputs of the national study should then be compared with the outputs of the Kehlbacher (2010) study (and other data) for the purposes of validation.

2. A research study should be commissioned by Defra to pilot the methodology that has been developed for the policy evaluation tool in the current study. This would test the technical and practical feasibility of applying the tool. This pilot need not construct a full-scale version of the tool and could be limited to one species (e.g. laying hens or broilers), but it should be sufficient to identify the practical merits and limitations of the tool in terms of its use for policy evaluation purposes.

3. There needs to be periodic collection of data from on-farm welfare assessments (e.g. using the Welfare Quality or adapted protocols) for the purposes of (i) deriving inputs into the policy evaluation tool; and (ii) informing expert stakeholders for the purpose of updating and revising the technical coefficients in the tool.

9. References

Alberini, A., Boyle, K. and Welsh, M. P. (2003) Analysis of contingent valuation data with multiple bids and response options allowing respondents to express uncertainty. *Journal of Environmental Economics and Management*, 45, 40–62.

Alfnes, F. and Rickertsen, K. (2003) European consumers' willingness to pay for US beef in experimental auction markets. *American Journal of Agricultural Economics*, 85, 396–405.

Alfnes, F. and Rickertsen, K. (2007) Extrapolating experimental-auction results using a stated choice survey. *European Journal of Agricultural Economics*, 34, 345–363.

Alfnes, F., Guttormsen, A., Steine, G. and Kolstad, K. (2005) *Consumers' willingness to pay for the color of salmon: a choice experiment with real economic incentives*. Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting, Providence, Rhode Island, 24-27 July 2005.

Anonymous (2010) Activists and animal welfare: quality verifications in the pork sector. Unpublished.

Association of Zoos and Aquariums (2009) *The accreditation standards and related policies*. 2008 ed. Association of Zoos and Aquariums, Silver Springs, MD. 60pp.

Assured Food Standards (2009) Personal communication with David Clarke, Chief Executive, 24 July.

Baltzer, K. (2004) Consumers' willingness to pay for food quality - the case of eggs. *Acta Agriculturae Skandinavica, Section C*, 1, 78–90.

Barber, J.C.E. (2009) Programmatic approaches to assessing and improving animal welfare in zoos and aquariums. *Zoo Biology,* 28, 519-530.

Barnett, J.L. and Hemsworth, P.H. (1990) The Validity of Physiological and Behavioural Measures of Animal Welfare. *Applied Animal Behaviour Science,* 25.

Bartussek, H. (1991) A Concept to Define Naturalness in Animal Production. In: *Proceedings of the International Conference on Alternatives in Animal Husbandry*, Witzenhausen, 22-25 July 1991. University of Kassel, 309-319.

Bartussek, H. (1995a) *Tiergerechtheitsindex für Legehennen*, TGI 35/L. Veröffentlichungen Nr. 25. Bundesanstalt für alpenländische Landwirtschaft (BAL) Gumpenstein, A 8952, Irdning.

Bartussek, H. (1995b) *Tiergerechtheitsindex für Mastschweine*, TGI 35/L. Stand Dezember 1995. Bundesanstalt für alpenländische Landwirtschaft (BAL) Gumpenstein, A 8952, Irdning.

Bartussek, H. (1996a) *Tiergerechtheitsindex für Rinder*, TGI 35/L. Stand Mai 1996. Bundesanstalt für alpenländische Landwirtschaft (BAL) Gumpenstein, A 8952, Irdning.

Bartussek, H. (1996b) *Tiergerechtheitsindex für Kälber*, TGI 35/L, Stand Mai 1996, Bundesanstalt für alpenländische Landwirtschaft (BAL) Gumpenstein, A 8952, Irdning.

Bartussek, H. (1999a) *Tiergerechtheitsindex für leere und tragende Sauen*. Stand Juni 1999. Bundesanstalt für alpenländische Landwirtschaft (BAL) Gumpenstein, A 8952, Irdning.

Bartussek, H. (1999b) A review of the animal needs index (ANI) for the assessment of animals` well-being in the housing systems for Austrian proprietary products and legislation. *Livestock Production Science*, 61, 179–192.

Baumans, V. (2005) Science-based assessment of animal welfare: laboratory animals. *Rev. sci. tech. Off. int. Epiz.,* 24, 503-514.

Bennett, R. M. (1995) The value of farm animal welfare. *Journal of Agricultural Economics*, 46, 46–60.

Bennett, R. M. (1996) Willingness-to-pay measures of public support for farm animal welfare legislation. *Veterinary Record*, 139, 320–321.

Bennett, R. M. (1997) Farm animal welfare and food policy. *Food Policy*, 22, 281–288.

Bennett, R. M., Anderson, J. and Blaney, R.J.P. (2002) Moral intensity and willingness to pay concerning farm animal welfare issues and the implications for agricultural policy. *Journal of Agricultural and Environmental Ethics*, 15, 187–202.

Bennett, R. M. and Blaney, R. (2002) Social consensus, moral intensity and willingness to pay to address a farm animal welfare issue. *Journal of Economic Psychology*, 23, 501–520.

Bennett, R. M. and Blaney, R. (2003) Estimating the benefits of farm animal welfare legislation using the contingent valuation method. *Agricultural Economics*, 29, 85–98.

Bennett, R. M., Broom, D. M., Henson, S., Blaney, R. and Harper, G. (2004) Assessment of the impact of government animal welfare policy on farm animal welfare in the UK. *Animal Welfare*, 13, 1–11.

Bennett, R. M. and Larson, D. (1996) Contingent valuation of the perceived benefits of farm animal welfare legislation: An exploratory survey. *Journal of Agricultural Economics*, 47, 224–235.

Bennett, R. M. and Tranter, R. B. (1998) The dilemma concerning choice of contingent valuation willingness-to-pay elicitation format. *Journal of Environmental Planning and Management*, 41, 253–257.

Botreau, R., Bonde, M., Butterworth, A., Perny, P., Bracke, M. B. M., Capdeville, J. and Veissier, I. (2007a). Aggregation of measures to produce an overall assessment of animal welfare. Part 1, A review of existing methods. *Animal,* 1, 1179–1187.

Botreau, R., Bracke, M.B.M., Perny, P., Butterworth, A., Capdeville, J., van Reenen, C.G. and Veissier, I. (2007b) Aggregation of measures to produce an overall assessment of animal welfare. Part 2, Analysis of constraints. *Animal,* 1, 1188-1197.

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

Bracke, M.B.M., Spruijt, B.M., Metz, J.H.M. and Schouten, W.G.P. (2002) Decision support system for overall welfare assessment in pregnant sows. A: Model structure and weighting procedure. *Journal of Animal Science*, 80, 1819-1834.

Brambell, R. (1965) *Report of the technical committee to enquire into the welfare of animals kept under intensive livestock husbandry systems.* London: HMSO.

Bryan, S. and Dolan, P. (2004) Discrete choice experiments in health economics. For better or for worse? *European Journal of Health Economics*, 5, 199-202.

Buller, H. and Morris, C. (2002) *Farm animal welfare aspects of the CAP*. London: RSPCA.

Buller, H. and Roe, E.J. (2009) *Constructing Quality: Negotiating Farm Animal Welfare in Food Assurance Schemes*. Proceedings of the Welfare Quality® Advisory Commitee Meeting, 24-25 September 2008, Copenhagen.

Burgess, D. and Hutchinson, W.G. (2005) Do people value the welfare of farm animals? *EuroChoices*, 4, 36–43.

Burgess, D., Hutchinson, W.G., McCallion, T., and Scarpa, R. (2003) *Investigating choice rationality in stated preference methods for enhanced farm animal welfare*. CSERGE Working Paper, ECM 03-02. University of East Anglia, Norwich.

Butterworth, A., Veissier, I., Manteca, J. X. and Blokhuis, H. J. (2008) Welfare Trade. European Union 15, *Public Service Review*, 18:54:39, 456-459.

Buzby, J. C., Fox, J. A., Ready, R. C., and Crutchfield, S. R. (1998) Measuring consumer benefits of food safety risk reduction. *Journal of Agricultural and Applied Economics*, 30, 69–82.

Callan, S. J. and Thomas, J. M. (2004) Environmental Economics and Management: Theory, Policy, and Applications. Thomson, South-Western, Ohio, USA.

Caritas Data (2009) *Top 3,000 Charities, 2008-09*. London: Caritas Data.

Carlsson, F., Frykblom, P. and Lagerkvist, C. J. (2005a) Consumer preferences for food product quality attributes from Swedish agriculture. *Ambio*, 34, 366–370.

Carlsson, F., Frykblom, P. and Lagerkvist, C. J. (2005b) Using cheap talk as a test of validity in choice experiments. *Economics Letters*, 89, 147–152.

Carlsson, F., Frykblom, P. and Lagerkvist, C. J. (2007a) Consumer willingness to pay for farm animal welfare: mobile abattoirs versus transportation to slaughter. *European Review of Agricultural Economics*, 34, 321–344.

Carlsson, F., Frykblom, P. and Lagerkvist, C. J. (2007b) Farm animal welfare – testing for market failure. *Journal of Agricultural and Applied Economics*, 39, 61–73.

Carlstead, K. and Brown, J.L. (2005) Relationships between patterns of fecal corticoid excretion and behavior, reproduction and environmental factors in captive black (*Diceros bicornis*) and white (*Ceratotherium simum*) rhinoceros. *Zoo Biology,* 24, 215-232.

Carson, R. (1985) *Three essays on contingent valuation (welfare economics, non-market goods, water quality)*. Department of Agricultural and Resource Economics, University of California, Berkeley.

Champ, P. A., Boyle, K. J. and Brown, T. C. (2003) *A Primer on Nonmarket Valuation*. London, Springer Verlag.

Charities Aid Foundation/National Council for Voluntary Organisations (2007) *UK Giving 2007*. London: CAF/NCVO.

Charities Aid Foundation/National Council for Voluntary Organisations (2009) *UK Giving 2009*. London: CAF/NCVO.

Charity Commission (2010) Income of registered animal charities.
http://www.charity-commission.gov.uk/Showcharity/RegisterofCharities/SearchMatchList.aspx

RegisteredCharity Number=0&SubsidiaryNumber=0

Dawkins, M.S. (1990) From an animal's point of view: motivation, fitness, and animal welfare. *Behavioural and Brain Sciences*, 13, 1-61.

Defra (2007) Animal Welfare Delivery Strategy. Defra, London. http://www.defra.gov.uk/foodfarm/policy/animalhealth/documents/awdelivery-strategy.pdf

Defra (2008) *Sustainable Farming and Food Strategy – indicator data sheet. Core indicator 5.2: Farm assurance schemes*. London, Defra.

DeShazo, J. R. and Fermo, G. (2002) Designing choice sets for stated preference methods: the effects of complexity on choice consistency. *Journal of Environmental Economics and Management*, 44, 123–143.

Diamond, P. A. and Hausman, J. A. (1994) Contingent valuation: Is some number better than no number? *Journal of Economic Perspectives*, 8, 45-64.

Dickinson, D. L., Hobbs, J. E. and Bailey, D. (2003) *A comparison of US and Canadian consumers' willingness to pay for red-meat traceability.* Paper presented at the American Agricultural Economics Association Annual Meeting, Montreal, Canada, 27-30 July.

Diener, E., Suh, E. M., Lucas, R. E. and Smith, H. L. (1999) Subjective well-being: three decades of progress. *Psychological Bulletin*, 125, 276–302.

Dolan, P. and Metcalfe, R. (2008) *Comparing willingness-to-pay and subjective well-being in the context of non-market goods*. CEP Discussion Paper No 890. Centre for Economic Performance, London School of Economics and Political Science.

Dolan, P. and Peasgood, T. (2006) *Valuing non-market goods: does subjective well-being offer a viable alternative to contingent valuation?* Unpublished manuscript. Tanaka Business School, Imperial College, London.

Ecos Consultancy (2009) *Encouraging philanthropic support to improve the welfare of farmed animals*. London, Ecos Consultancy Ltd.

Enneking, U. (2004) Willingness to pay for safety improvements in the German meat sector: the case of the Q&S label. *European Review of Agricultural Economics*, 31, 205–223.

European Commission (2007) Attitudes of EU citizens towards animal welfare. *Special Eurobarometer*, 270.

European Parliament (2001) *Communication from the Commission to the Council and the European Parliament on the welfare of intensively kept pigs in particularly taking into account the welfare of sows reared in varying degrees of confinement and in groups*. COM(2001) 20 final, 16.01.2001. http://europa.eu.int/eur-lex/en/com/pdf/2001/com2001_0020en01.pdf

Faure, J.M., Bessei, W. and Jones, R.B. (2003) Direct selection for improvement of animal well-being. In: Muir, W. and Aggrey, S. (eds.) *Poultry Breeding and Biotechnology*. CAB International, 221-245.

FAWC (2010) *Farm Animal Welfare in Great Britain: Past, Present and Future*. Farm Animal Welfare Council, London (www.fawc.org.uk).

Ferrer-i Carbonell, A. (2002) *Subjective questions to measure welfare and well-being: a survey*. Tinbergen Institute Discussion Paper No. 02-020/3. Faculty of Economics and Econometrics and SEO, University of Amsterdam.

Frey, B. S., Luechinger, S. and Stutzer, A. (2004) *Valuing public goods: the life satisfaction approach*. Working Paper No. 184. Institute for Empirical Research in Economics, University of Zurich.

Geers, R., Petersen, B., Huysmans, K., Knura-Deszczka, S., De Becker, M., Gymnich, S., Henot, D., Hiss, S. and Sauerwein, H. (2003) On-Farm Monitoring of Pig Welfare by Assessment of Housing, Management, Health Records and Plasma Haptoglobin. *Animal Welfare,* 12, 643-647.

Glass, C. A., Hutchinson, W. G. and Beattie, V. E. (2005) Measuring the value to the public of pig welfare improvements: a contingent valuation approach. *Animal Welfare*, 14, 61-69.

Gracia, A., Loureiro, M. and Nayga, R. M. (2009) *Do consumers value animal welfare products?* Paper presented at a workshop on valuation methods in agro-food and environment economics, experimental auctions: theoretical background and empirical applications. Barcelona, 2-3 July.

Hamilton, S. F., Sunding, D. L. and Zilberman, D. (2003) Public goods and the value of product quality regulations: the case of food safety. *Journal of Public Economics*, 87, 799-817.

Hanemann, W. M. (1984) Welfare evaluations in contingent valuation experiments with discrete response. *American Agricultural Economics Association*, 66, 332-341.

Hanemann, W. M. (1985) Some issues in continuous and discrete response contingent valuation studies. *Northeastern Journal of Agricultural Economics*, 14, 5-13.

Hanley, N., Shogren, J. and Wright, B. (2001) *Introduction to Environmental Economics*. Oxford University Press.

Harrison, G. W. and Rustrom, E. E. (2008) Attribute based methods. In: Plott, C. R. and Smith, V. L. (Eds*) Handbook of experimental economics results.* Amsterdam, Elsevier.

Henson, S. J. and Beard, N. F. (1994) Practical techniques for valuing farm animal welfare. In: Bennett, R. M. (Ed.) *Valuing farm animal welfare*. Proceedings of a workshop held at the University of Reading, 30 September 1993. University of Reading.

Hitzhusen, F. J. (2007) *Economic valuation of river systems*. Cheltenham, Edward Elgar Publishing.

Home Office (2001) *Review of the 'Ethical Review Process' in establishments designated under the Animals (Scientific Procedures) Act 1986.* London, TSO.

House of Lords (2002) *Report of the Select Committee on animals in scientific procedures*. HL Paper 150-1. London, TSO.

Hurnik, J.F. (1990) World's poultry science association invited lecture: Animal welfare: ethical aspects and practical considerations. *Poultry Science,* 69, 1827-1834.

ILAR (2000) *Institute for Laboratory Animal Research Journal,* 41(2). http://delsold.nas.edu/ilar_n/ilarjournal/41_2/index.shtml

Institute of Grocery Distribution (2010a*) The Growth of Ethical Shopping*. http://www.igd.com/print.asp?pid=1&pflid=1&plid=5&pcid=291

Institute of Grocery Distribution (2010b) *Ethical and Sustainable Shopping*. London, IGD.

Jaeger, S. R., Lusk, J. L., House, L. O., Valli, C., Moore, M., Morrow, B., and Traill, W. B. (2004) The use of non-hypothetical experimental food markets for measuring the acceptance of genetically modified foods. *Food Quality and Preference*, 15, 701-714.

Johnson, R. F. and Desvousges, W. H. (1997) Estimating stated preferences with rated-pair data: Environmental, health, and employment effects of energy programs. *Journal of Environmental Economics and Management*, 34, 79-99.

Jones, R.B. (2001) Environmental enrichment for poultry welfare. In: Wathes, C.M. (ed.) *Integrated Management Systems for Livestock*. British Society for Animal Science, Occasional Publication, No 28, 125-131.

Kahneman, D. and Krueger, A. B. (2006) Developments in the measurement of subjective well-being. *Journal of Economic Perspectives*, 20, 3-24.

Keeling, L. and Svedberg, J. (1999) Legislation banning conventional battery cages in Sweden and a subsequent phase-out programme. In: Kunisch, M. and Eckel, H. (eds.) *Proceedings of the Congress 'Regulation of Animal Production in Europe'*, 73-78.

Keeling, L. and Bock, B. (2007) *Turning welfare principles into practice*. Proceeding of the 2nd Welfare Quality Stakeholder Conference. 3-4 May 2007, Berlin, 25-28.

Kehlbacher, A. (2010) *Willingness-to-pay for Animal Welfare in Livestock Production - An Application of the Choice Experiment and Contingent Valuation Methods in Great Britain*. PhD thesis, Department of Agricultural and Food Economics, University of Reading,

Kjær, T. (2005) *A review of the discrete choice experiment - with emphasis on its application in health care*. Health Economics Paper 2005, 1, University of Southern Denmark.

Kjaernes, U. and Larvik, R. (eds.) (2007) *Attitudes of Consumers, Retailers and Producers to Farm Animal Welfare*. Welfare Quality Report 2. Cardiff University Press.

Krueger, A. B. and Schkade, D. A. (2007) *The reliability of subjective well-being measures*. NBER Working Paper No. 13027. University of California, San Diego.

Lagerkvist, C. J., Carlsson, F. and Viske, D. (2006) Swedish consumer preferences for animal welfare and biotech: a choice experiment. *AgBio Forum,* 9, 51-58.

Lagerkvist, C. J. and Hess, S. (2011) A meta-analysis of consumer willingness to pay for farm animal welfare. *European Review of Agricultural Economics*, 38, 55-78.

Lampkin, N., Measures, M. and Padel, S. (Eds) (2008) *2009 Organic Farm Management Handbook*. Aberystwyth University and Organic Research Centre, Hampstead Marshall.

Lancaster, K. J. (1966) A new approach to consumer theory. *The Journal of Political Economy*, 74, 132-157.

Leach, M.C. and Main, D.C.J (2008) An assessment of laboratory mouse welfare in UK animal units. *Animal Welfare*, 17, 171-187.

Leach, M.C, Thornton, P.D. and Main, D.C.J (2008) Identification of appropriate measures for the assessment of laboratory mouse welfare. *Animal Welfare*, 17, 161-170.

Leon, C. J. and Vazquez-Polo, F. J. (1998) A Bayesian approach to double bounded contingent valuation. *Environmental and Resource Economics*, 11, 197-215.

Liljenstolpe, C. (2008) Evaluating animal welfare with choice experiments: an application to Swedish pig production. *Agribusiness,* 24, 67-84.

List, J. A., Sinha, P. and Taylor, M. H. (2006) Using choice experiments to value non-market goods and services: Evidence from field experiments. *Advances in Economic Analysis and Policy*, 6(2): Article 2.

Lockwood, M. (1996) Non-compensatory preference structures in non-market valuation of natural area policy. *Australian Journal of Agricultural Economics,* 40, 85-101.

Loomis, J. and Helfand, G. (2003) *Environmental Policy Analysis for Decision Making*. New York, Kluwer Academic Publishers.

Louviere, J., Hensher, D. A. and Swait, J. (2000) *Stated Choice Methods - Analysis and Applications.* Cambridge University Press.

Lusk, J. L. and Fox, J. A. (2002) Consumer demand for mandatory labelling of beef from cattle administered growth hormones of fed genetically modified corn. *Journal of Agricultural and Applied Economics*, 34, 27-38.

Lusk, J. L., Nilsson, T. and Foster, K. (2007) Public preferences and private choices: effect of altruism and free riding on demand for environmentally certified pork. *Environmental and Resource Economics*, 36, 499–521.

Lusk, J. L., Roosen, J. and Fox, J. A. (2003) Demand for beef from cattle administered growth hormones or fed genetically modified corn: a comparison of consumers in France, Germany, the United Kingdom and the United States. *American Journal of Agricultural Economics*, 85, 16-29.

Lusk, J. L. and Schroeder, T. C. (2004) Are choice experiments incentive compatible? A test with quality differentiated beef steaks. *American Journal of Agricultural Economics*, 86, 467–482.

Main, D.C.J., Webster, J. and Green, L.E. (2001) Animal Welfare Assessment in Farm Assurance Schemes. *Acta Agriculturae Scandinavica, Section A*, *Animal Science Supplementum,* 30, 108-113.

Mayfield, L.E., Bennett, R.M., Tranter, R.B. and Wooldridge, M.J. (2007) Consumption of welfare-friendly food products in Great Britain, Italy and Sweden, and how it may be influenced by consumer attitudes to, and behaviour towards, animal welfare attributes. *International Journal of the Sociology of Agriculture and Food,* 15(3), 59-73.

McInerney, J. (1994) Animal welfare: an economic perspective. In: Bennett, R. M. (ed.) *Valuing farm animal welfare*. Proceedings of a workshop held at the University of Reading, 30 September 1993. University of Reading.

McInerney, J. (2004) *Animal welfare, economics and policy*. Report on a study undertaken for the Farm and Animal Health Economics Division of Defra. http://www.defra.gov.uk/evidence/economics/foodfarm/reports

McVittie, A., Moran, D. and Nevison, I. (2006) *Public preferences for broiler chicken welfare: evidence from stated preference studies*. Land Economy Working Paper Series, Number 3. Department of Land Economy, University of Cambridge.

McVittie, A., Moran, D., Sanilands, V. and Sparks, N. (2005) *Estimating non-market benefits of reduced stocking density and other welfare increasing measures for meat chickens in England*. Final report to Defra, SAC Project no. AW0236.

Melton, B. E., Huffman, W. E., Shogren, J. F. and Fox, J. A. (1996) Consumer preferences for fresh food items with multiple quality attributes: evidence from an experimental auction of pork chops. *Agricultural and Applied Economics Association*, 78, 916-923.

Miele, M. and Evans, A. (2005) European consumers' views about farm animal welfare. In: Butterworth, A. (Ed) *Science and society improving animal welfare*, Welfare Quality conference proceedings. 17-18 November 2005, 9-25. University of Cardiff.

Mitchell, R. C. and Carson, R. T. (1989) *Using surveys to value public goods: the contingent valuation method*. Resources for the Future, Washington DC.

Moran, D. and McVittie, A. (2008) Estimation of the value the public places on regulations to improve broiler welfare. *Animal Welfare*, 17, 43-52.

Moreira, N., Brown, J.L., Moraes, W., Swanson, W.F. and Monteiro-Filho, E.L.A. (2007) Effect of housing and environmental enrichment on adrenocortical activity, behavior and reproductive cyclicity in the female Tigrina (*Leopradus tigrinus*) and Margay (*Leopardus wiedii*). *Zoo Biology,* 26, 441-460.

Morkbak, R. N., Christens, T. and Gyrd-Hansen, D. (2010) Consumer preferences for safety characteristics in pork. *British Food Journal*, 112, 775-791.

Napolitano, F. (2009) Meat liking, animal welfare and consumer willingness to pay. *Italian Journal of Animal Science,* 8, 469-476.

Napolitano, F., Pacelli, C., Girolami, A. and Braghieri, A. (2008) Effect of information about animal welfare on consumer willingness to pay for yoghurt. *Journal of Dairy Science*, 91, 910-917.

Nocella, G., Hubbard, L. and Scarpa, R. (2010) Farm animal welfare, consumer willingness to pay, and trust: Results of a cross-national survey. *Applied Economic Perspectives and Policy*, 32, 275-297.

Norwood, F. B. and Lusk, J. L. (2008) *A calibrated auction-conjoint valuation method: valuing pork and eggs produced under differing animal welfare conditions*. Working Paper. Department of Agricultural Economics, Oklahoma State University.

OECD (2002) *OECD test guideline 401 was deleted in 2002: a major step in animal welfare: OECD reached agreement on the abolishment of the LD50 acute toxicity test.* http://www.oecd.org/document/52/0.2340.en_2649_34377_2752116_1_1_1_1.00.html

Olesen, I., Alfnes, F., Rora, B. M. and Kostad, K. (2010) Eliciting consumers' willingness to pay for organic and welfare-labelled salmon in a non-hypothetical choice experiment. *Livestock Science*, 127, 218–226.

Pearce, D. and Barbier, E. B. (2000). *Blueprint for a sustainable economy*. London, Earthscan Publications.

Rolfe, J. (1999) Ethical rules and the demand for free range eggs. *Economic Analysis and Policy*, 29, 187–206.

Rosen, S. (1974) Hedonic prices and implicit markets: product differentiation in pure competition. *The Journal of Political Economy*, 82, 34–55.

RSPCA's Freedom Food Ltd. (2009) Personal communication with Caroline Gauntlett, Consumer Marketing Manager, 14 August.

Rutter, S.M. (1998) Assessing the welfare of intensive and extensive livestock. In: *Proceedings of the Workshop Pasture Ecology and Animal Intake*, 1-9.

Shepherdson, D.J, Carlstead, K.C. and Wielebnowski, N. (2004) Cross-institutional assessment of stress responses in zoo animals using longitudinal monitoring of fecal corticoids and behavior. *Animal Welfare,* 13, S105-113.

Sørensen, P. (2001) Breeding strategies in poultry for genetic adaptation to the organic environment. *Proceedings, 4th NAHWOA Workshop*. Wageningen, Netherlands. 9pp.

Spoolder, H., De Rosa, G., Horning, B., Waiblinger, S. and Wemelsfelder, F. (2003) Integrating

Parameters to Assess on-Farm Welfare. *Animal Welfare,* 12, 529-534.

Stutzer, A. and Frey, B. S. (2010) *Recent advances in the economics of individual subjective well-being.* Discussion Paper No. 4850. Institute for the Study of Labor, Faculty of Business and Economics, University of Basel.

Taylor, N. and Signal, T. D. (2009) Willingness to pay: Australian consumers and 'On the farm' welfare. *Journal of Applied Animal Welfare Science*, 12, 345-359.

The Co-operative Bank (2009) Ten years of ethical consumerism 1999-2008. http://www.goodwithmoney.co.uk/assets/Ethical-Consumerism-Report-2009.pdf

The Co-operative Bank (2010) *Ethical Consumerism Report 2010. Ethical shopping through the downturn.* Manchester: The Co-operative Bank.

Tonsor, G. T., Olynk, N. and Wolf, C. (2009) Consumer preferences for animal welfare attributes: the case of gestation crates. *Journal of Agricultural and Applied Economics*, 41, 713–730.

Tonsor, G. T., Schroeder, T. C., Fox, J. A. and Biere, A. (2005) European preferences for beef steak attributes. *Journal of Agricultural and Resource Economics*, 30, 367-380.

Tversky, A. and Kahneman, D. (1974) Judgement under uncertainty: heuristics and biases. *Science*, 185, 1124-1131.

van Hoorn, A. (2007) *A short introduction to subjective well-being: its measurement, correlates and policy uses*. Paper prepared for the conference: 'Is happiness measurable and what do those measures mean for policy?' held in Rome, 2-3 April 2007.

Veissier, I., Butterworth, A., Bock, B. and Roe, E. (2008) European approaches to ensure good animal welfare. *Applied Animal Behaviour Science,* 113, 279-297.

Veisten, K. and Navrud, S. (2006) Contingent valuation and actual payment for voluntarily provided passive-use values: assessing the effect of an induced truth-telling mechanism and elicitation formats. *Applied Economics*, 38, 735-756.

Venkatachalam, L. (2004) The contingent valuation method: a review. *Environmental Impact Assessment Review*, 24, 89-124.

Von Borell, E., Bockisch, F.J., Buscher, W., Hoy, S., Krieter, J., Muller, C., Parvizi, N., Richter, T., Rudovsky, A., Sundrum, A. and Van Den Weghe, H. (2001) Critical Control Points for on-Farm Assessment of Pig Housing. *Livestock Production Science,* 72, 177-184.

Webster, J. (1997) Applied ethology: what use is it to animal welfare? *Advances in Ethology*, 32, Supplements to Ethology, 10.

Wielebnowski, N.C., Fletchall, N., Carlstead, K., Busso, J.M. and Brown, J.L. (2002) Non-invasive Assessment of Adrenal Activity Associated with Husbandry and Behavioral Factors in the North American Clouded Leopard Population. *Zoo Biology,* 21, 77-98.

Appendix A. Additional information and examples of animal based assessment data from laboratory, zoo and farm animals

A1. Some resource and animal based measures of welfare

Table A1. Resource and animal based measures used to assess adequacy of the housing environment for laboratory animals (mice) and the welfare status of these animals.

Measurements	
Cage specifications:	**Appearance:**
Stocking density	Physical damage
Group size	Posture
Environment:	Starey coat
Cage cleanliness	Alertness
Use of mouse resources:	Barbering
Use of nesting material	Discharges
Use of a shelter	Body score
Use of gnawing material	Openness of eyes
Use of a wheel	Chromodacryorrhea
Behaviour:	Hair loss
Behavioural repertoire	Wall hugging
Gait	Sunken abdomen
Quantity of cage gnawing	Pinched face
Tail position	Skin colour
Provoked responses:	**Individual health:**
Response to observer	Obvious disease signs
Response to novel object	Blood/saliva in bedding
Response to auditory stimulus	Weight gain
Vocalisation:	Faecal/urine output
Level of ultrasonic vocalisation	Nervous problems
Level of audible vocalisation	Respiration rate
	Respiration type
	Faecal glucocorticoid level

Source: Leach and Main, 2008; Leach *et al.*, 2008.

Table A2. A list of 'resources' which may be assessed in laboratory animal studies - these measures will not directly reflect the animals welfare, but alterations in these resources may have very significant impacts on the animal – which may (or sometimes may not) be measured using ABMs.

Measurements	
Cage specifications:	**Husbandry:**
Floor type	Handling competence
Presence of substrate	**Food/Water:**
Cage dimensions	Type of food
Type of ventilation	Method used to provide water
Cage material	**Staffing:**
Cage safety	Ease of observation of animals
Cage opaqueness	**Presence of resources:**
Environment:	Presence of nesting material
Light intensity	Presence of a shelter
Ultrasonic noise level	Presence of gnawing material
Ammonia level	Presence of visual barriers
Temperature	Presence of a wheel
Audible noise level	**Presence of others:**
Humidity	Presence of other species
CO2 levels	Presence of other mouse cages in animal room
Particulate levels	

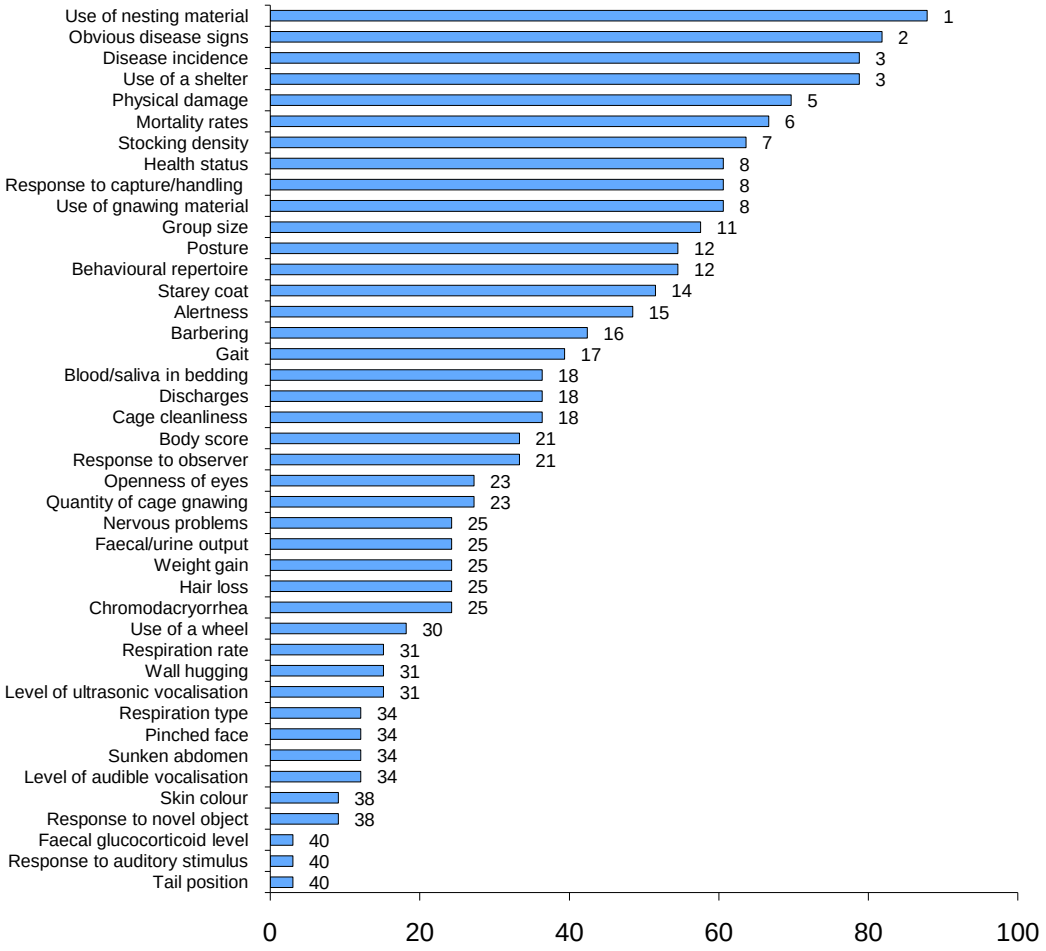
Source: Leach and Main, 2008; Leach *et al*., 2008.

Table A3. The animal-based outcome and resource input measures that are recorded in questionnaire form (i.e. those which require information from an animal carer and may not be immediately apparent from short term observation).

Measurements	Resource input	Animal-based outcome
Provoked responses:		
Response to capture/handling		✓
Environment:		
Photoperiod times	✓	
Frequency of lights 'on' in sleep phase	✓	
Source of light in animal rooms	✓	
Husbandry:		
Frequency of re-grouping	✓	
Euthanasia method	✓	
Cage cleaning frequency	✓	
Cage cleaning method	✓	
Quality of facilities for sick/injured	✓	
Euthanasia of sick/injured	✓	
Weaning age	✓	
Euthanasia with conspecifics present	✓	
Method used to identify animals	✓	
Interference with animals in sleep phase	✓	
Type of transport within an establishment	✓	
Food/Water:		
Occurrence of flooding	✓	
Presence of floor food	✓	
Staffing:		
Staff attitude	✓	
Inspection of animals	✓	
Staff training	✓	
Staffing levels	✓	
Availability of welfare information	✓	
Availability of breeding records	✓	
Membership of an accreditation scheme	✓	
Presence of resources:		
Presence of background music	✓	
Presence of others:		
Animal room activity	✓	
Unit Health:		
Disease incidence		✓
Mortality rates		✓
Health status		✓
Health screening process	✓	
Welfare considered as by ethical review	✓	
Disease limitation methods	✓	
Hazard assessment scheme	✓	

Source: Leach and Main, 2008; Leach *et al*., 2008.

Figure A1**.** The percentage number of the 29 participants that chose each animal-based outcome measure in descending rank order (the actual ranks are shown at the end of each column).



Source: Leach and Main, 2008; Leach *et al*., 2008.

Table A4. Reported and observed prevalence of aggression, stereotypy and climbing behaviour.

Aggression	Overall	1.7%
	Non-breeding	1.9%
	Breeding	0.6%
Stereotypy	Overall	12.2%
	Non-breeding	13.7%
	Breeding	1.2%
Climbing	Overall	22.8%
	Non-breeding	25.6%
	Breeding	3.1%

Source: Leach and Main, 2008; Leach *et al*., 2008.

Table A5. Prevalence of potentially abnormal behaviours (considered abnormal and/or detrimental) reported by the animal units.

Barbering	Reported	2.4%
	Observed	4.2%
Physical damage	Reported	2.4%

Source: Leach and Main, 2008; Leach *et al*., 2008.

A2. Information on the Animal Needs Index (ANI)

The first well-described systematic assessment system based (partly) on animal outcomes was the Austrian 'Animal Needs Index' (ANI), or 'Tiergerechtheitsindex' (TGI) concept which was first described as in Styrian (an Austrian state) Animal Welfare Regulation in 1985 and was further developed and finally 'published' in 1988 in Naturgemasse Viehwirtschaft. In 1990 the system was first applied in a trial on 118 organic farms and, in 1991, the 'full version' (ANI35L - *'L' for Long*) was published for cattle and pigs, and then adopted into Vorarlberg (Austrian animal welfare law) for cattle in 1992 (Bartussek 1991, 1995a, 1995b, 1996a, 1996b, 1999a, 1999b). Between 1995 and 1999 - it was broadly applied in Austria.

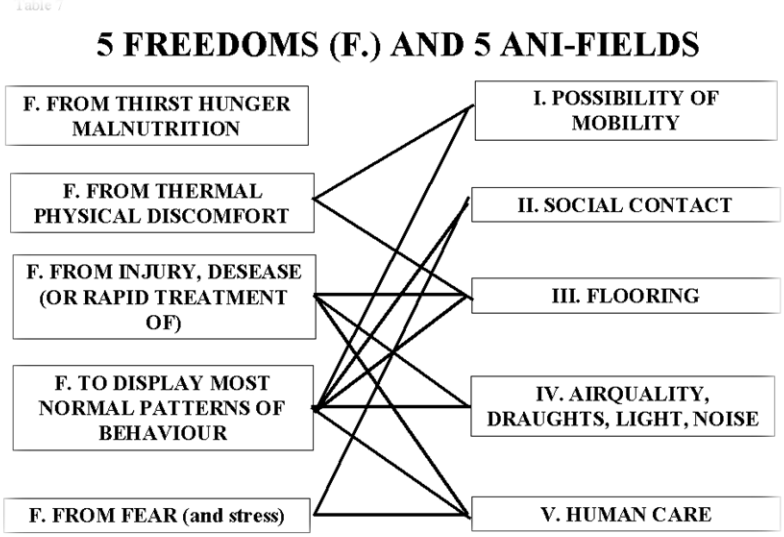
The ANI had the following aims:

(i) to meet the market demand and regulatory need for an assessment tool which could be used for all species production systems, methods and locations and used directly on farms; and

(ii) to include a grading of results to allow different standards of housing conditions with respect to the well-being of animals, with the aim of improving animal welfare step by step.

The ANI bases its approach to animal welfare assessment on the Five Freedoms (Brambell, 1965) as a basis for general principles of animal well-being. Four of the areas are 'freedoms' *from* something that causes discomfort, suffering or pain, and the remaining area concerns a freedom *to do* according to the *behavioural needs and wants* of the animal (see Table A6).

Table A6. ANI 'fields' in relation to the Five Freedoms.



Source: Bartussek (1996b).

Bartussek (1991) stated at the time of introduction of the ANI that: 'The real meaning of these Five Freedoms for the animals can be approached by innumerable physical, anatomical, physiological, pathological and ethological indicators. Today's knowledge about them fills libraries. To make an assessment tool practical, one must greatly reduce the number of indicators to be used in the system'.

The ANI principally considers five aspects of the animal's environment:

(i) the potential for mobility;
(ii) social contact;
(iii) the condition of the flooring for lying, standing and walking;
(iv) the housing climate (ventilation, light and noise);
(v) the intensity or quality of human care.

The principles of the ANI35 were to: assess the housing conditions using a range of 'measures' each with its own numeric 'point score'; to scale the findings to give more points for better housing condition, summate the scores to give an ANI index value (a combined score); to set 'minimum requirements' to prevent scores below an unacceptable level from being tolerated; to rank the scores into welfare categories based on the index scores; to ensure that the system worked across the entire range of husbandry systems; and to 'negotiate' with the parties involved (to achieve positive management actions based on the scores resulting from the assessment).

Within each field, several species-specific criteria were graded by the use of 'points'. Conditions considered to give animals appropriate opportunity to satisfy their behavioural needs, or to improve their welfare, were awarded more points. The overall sum of the points gives the ANI value. Poor conditions in one area could be compensated by 'better' conditions within another field. However, a minimal critical condition was described which, if not fulfilled, meant that an ANI score could not be given. This 'proviso' or 'veto' was considered an important part of the ANI as it was aimed at preventing very poor findings from being tolerated due to the compensation that could take place due to high scores (better welfare states) in other measures. Effectively, this 'veto rule' prevented the averaging of scores to cancel out very poor welfare scores – and so prevented the system from permitting a farm which had elements of poor welfare from receiving an acceptable overall score. Table A7 shows, as an example, the score scale for laying hens used in the ANI.

Six overall welfare categories were proposed for the range of ANI scores (Table A8). The last column characterises the five categories above 'not sufficient with respect to welfare' by an increasing number of symbols (e.g. stars, actually small 'animal heads' in the original).

Tables A9 and A10 give an overview of the broad application of the ANI assessment tool in Austria between 1992 and 1999.

Table A7. The criteria of the ANI35L for laying hens (1995) showing the range of potential scores for each measure as well as the overall minimum and maximum sum of scores.

Fields of influence	Characteristics	Points: Min. - Max.
I. Possibility of movement	floor area per animal	-0.5 to 1.5
	% of floor littered for scraping	-0.5 to 1.5
	elevated perches available	0 to 1.0
	area of forcourt per animal	0 to 1.0
	days of outside exercise/year	0 to 1.5
	area of grassland per animal	0 to 1.5
	max. distance of grassland	-0.5 to 1.5
II. Social contact	animals per separate group	0 to 1.5
	floor area per animal	-0.5 to 1.0
	availability of important outfit	-0.5 to 1.5
	elevated perches available	0 to 1.0
	cocks within herd	0 to 1.0
	width of exit hole	-0.5 to 1.0
	mean distance to exit hole	-0.5 to 1.0
	outfit of forcourt	0 to 1.0
III. Quality of flooring	perch length per animal	-0.5 to 1.0
	quality of perches	0 to 1.0
	type of cover of dunging pit	0 to 1.0
	type/amount of litter (scraping area)	0 to 1.0
	condition of litter	-0.5 to 1.5
	type of flooring in next area	-0.5 to 1.5
	floor condition – forcourt	-0.5 to 1.0
	condition of grassland	-0.5 to 1.0
IV. Climatization	light within stable	-0.5 to 1.5
(light, ventilation, noise)	air quality within stable	-0.5 to 1.5
	draughts in resting area	-0.5 to 1.0
	technical noise	-0.5 to 1.0
	days of outside exercise/year	0 to 1.5
	hours of outside exercise/day	0 to 1.5
	shadow on grassland	-0.5 to 1.5
V. Care of stockman	cleanliness of housing	-0.5 to 1.5
	state of techn. Equipment	-0.5 to 1.5
	carcasses within stable	-0.5 to 1.0
	condition of plumage	-0.5 to 1.5
	condition of skin	-0.5 to 1.5
	recording of relevant data	0 to 1.0
	animal health	-0.5 to 1.5
Sum of points	absolute 33.5	-11.5 to 45.0

Source: Bartussek, 1995a.

Table A8. ANI welfare categories for the ANI35L-system.

Sum of ANI points	Names of categories of housing conditions with respect to welfare	Symbol
< 11	Not suitable	No label
11-16	Scarcely suitable	*
16.5-21	Little suitable (mediocre)	**
21.5 -24	Fairly suitable	***
24.5-28	Suitable	****
> 28	Very suitable	*****

Source: Bartussek, 1995a.

Table A9. History of the application of the ANI in Austria between 1995 and 1999.

APPLICATION OF ANI 35 L

1995 **official introduction of ANI 35 L/1995 (cattle, laying hens and fattening pigs) in controlling organic farms**
minimum requirement:
21 ANI points for existing buildings > (fairly suitable)
> 24 ANI points for new or rebuilt houses > (suitable)

1996 **official introduction of ANI 35 L/1996-calves in organic farming; ANI 35 L for cattle amended; minimum requirements (for proviso clause) for dairy cows defined**

1997 **official Austrian proposal of ANI for supplementation of EU-regulation 2092/91 for minimum requirements for tied housing of cows (not accepted); 20.000 farms controlled**

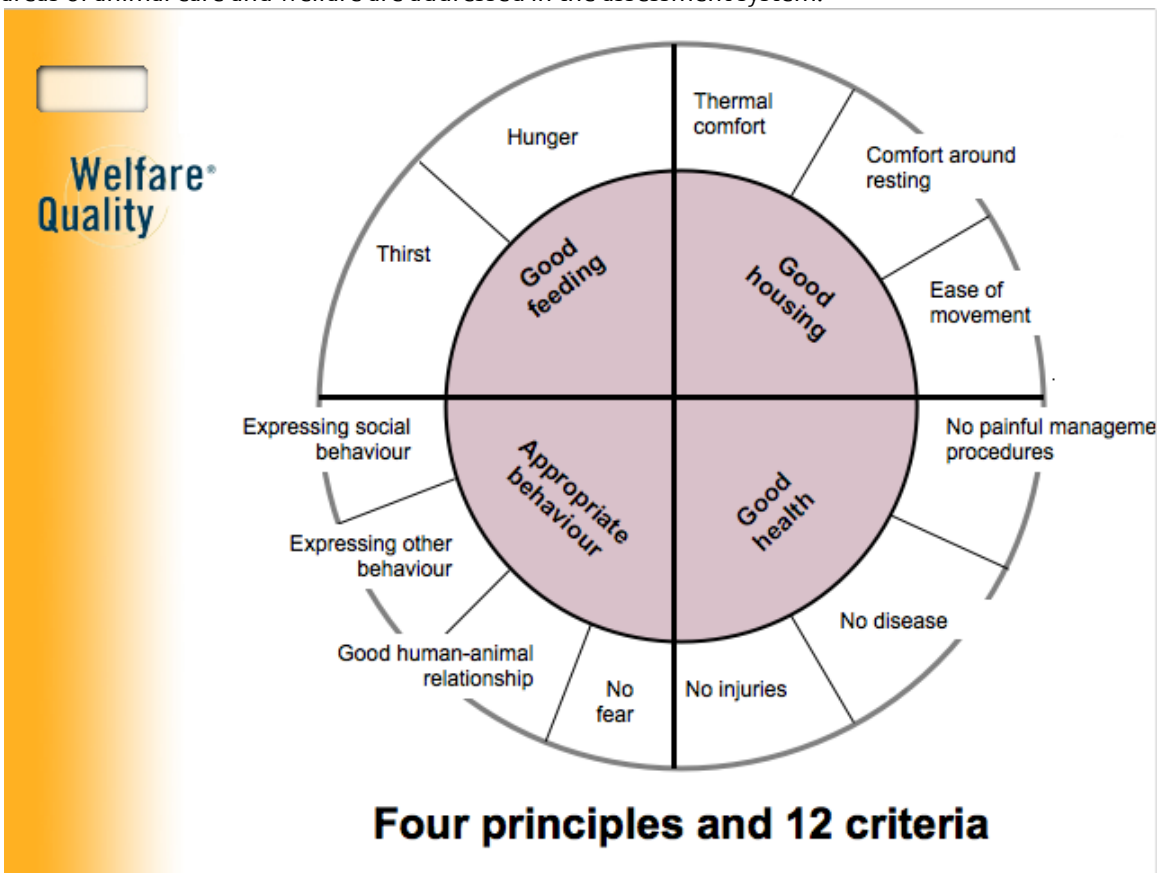
1999 **introduction of ANI 35 L/1996-sows; publication of all ANI 35 L and supplements by Codex Alimentarius Commission; definition of sufficient welfare for existing stables, tied housing for cows during transition period until 2010, for small holdings also thereafter**

Table A10. History of the application of the ANI in Austrian Legislation between 1992 and 1997.

APPLICATION OF ANI 35 L	
1992	**execution of animal welfare law of Vorarlberg (short version, ANI 5)**
1993-1995	**agreement of federal provinces about animal welfare standards for agricultural animals: 5 ANI fields of influence officially introduced**
1997	**Salzburg Agricultural Animal Protection Law (LGBI 76, 1997)(7,28,31): ANI 35 L by regulation of government (not issued yet)**
1997	**Tirol Animal Protection Regulation (LGBI 80/1997), short version ANI 5 (6, Annex 10)**

A3. Additional information on the Welfare quality assessment system

Figure A2. The 'wheel' of measures proposed by the Welfare Quality project which aim to cover a wide range of animal areas in a systematic way and so to provide a reasonable assurance that many areas of animal care and welfare are addressed in the assessment system.



The analysis of the inter-observer agreement between experts for each parameter can be carried out using the Kendall's coefficient of concordance W that measures the agreement between several assessors. Kendall W coefficients may range from 0 to 1. For 'an important category that is difficult to measure, a rough guideline for acceptability might be a correlation coefficient of at least 0.7'. Reliability testing for dichotomous variables (0/1) can be analyzed as an index of concordance, also called the observed proportion of agreement (PO). Table A11 summarises the list of parameters tested.

Table A11. List of parameters tested including test statistics.

Parameters	**Observations**	**Measure of agreement**
Avoidance distance at the feeding rail (ADF)	Live observations	Kendall's W
Lying down sequences duration	Video clips	Kendall's W
Herd scan	Live observations	Kendall's W
Continuous behaviour observations	Live observations, video clips	Kendall's W
Individual clinical scoring	Live observations	PO Proportion of agreement

A4. Some examples/descriptions of the way in which measures are made in the Welfare Quality tool

Appropriate behaviour (11) - avoidance distance at the feeding rail

Avoidance distance at the feeding rail (ADF) is a measure which evaluates the human-animal relationship on a given farm. This test is carried out in animals heavier than 200 kg. The observer places him/herself in front of the animal at a starting distance of 3 m, in order to prevent a first reaction of 'avoidance'. The head of the tested animal has to be completely outside the feeding rail, and the animal has to be alert and aware of the observer's presence. If an animal is not obviously attentive, but also not clearly distracted, it is tested. The assessor slowly approaches the animal with the arm reaching out towards the animal at an angle of approximately 45° to the body. During the approach, the back of the assessor's hand is directed towards the animal. The assessor never stares into the animal's eyes, but rather looks at the muzzle and continues to walk towards the animal until it shows signs of withdrawal or until the muzzle could be touched. As soon as the animal moves back, turns its head to one side or pulls back its head, the avoidance distance is estimated (= distance between the hand and the muzzle at the moment of withdrawal). Recordings of the avoidance distance are estimated in 10 cm intervals (from 300 cm to 10 cm). If an animal can be touched, the avoidance distance is noted as 0 cm. Neighbouring animals that react to an animal being

tested are tested later. In order to reduce the risk of disturbance every second animal is evaluated if the animals are 'lined up' at the feeding rail. If the space in front of the animals is less than 3 m, the test is carried out with the assessor positioned at an angle of maximum 45° to the feeding rack. The starting point is then at a distance of 3.5 meters from the rack. If a distance of 3.5 meters is not possible, the assessment is made from the maximum distance possible, which is written down on the recording sheet. Great care is taken that the observer's hand is always the body part closest to the animals during the approach, because sometimes animals react to the knee or the foot first when they are closer than the hand. Animals that appear obviously distracted are not tested. In order to be sure of the animal's withdrawal reactions, the assessor stops only when the animal shows clear signs of avoidance.

Behavioural observations (9 and 10)

1. Herd scan

During these scans both the number of animals that are 'lying and ruminating' and the number of animals that are 'standing and ruminating' are recorded per pen (up to a maximum of 15 pens). Herd scans are only applied to animals heavier than 350 kg and every animal has to be observed for at least 3 seconds for signs of rumination. Depending on the number of pens sampled and the respective duration of continuous behaviour observations, herd scans are carried out approximately every 30 minutes. The first scans are carried out before the beginning of behaviour observations and the last one at the end of the behaviour observations. The herd scan records the number of animals lying and ruminating, excluding those standing.

2. Segment scan

Before starting and after finishing the continuous behaviour observation in a pen, a 'unit scan' is performed, in order to record the number of feeding, lying and standing animals. 'Lying' animals are those that, during the segment scan, are resting on the belly and hind quarters or sitting dog-like on their hind quarters. 'Feeding/drinking' animals are those situated in the feed alley (feeding place) with their heads through the feeding rail or the neck rail, plus animals that are drinking. Animals neither lying nor feeding/drinking are considered to be 'standing'.

3. Continuous behaviour observations

Continuous behaviour sampling and one-zero sampling are carried out in order to quantify the animals social, resting and other behaviour. To ensure an ideal view of the pen, the observer is placed outside the animal area in the place that allows the least obstructed view. Interactions between animals situated in different parts of the pen are recorded if the animal being assessed is located in the area under observation. During the continuous behaviour observations, the occurrence of specific behaviours is recorded.

‘Head butt’ interaction involving physical contact where the assessed animal is butting, hitting, thrusting, striking or pushing the receiver with its forehead, horns or horn base with a forceful movement; the receiver does not give up its present position.

‘Displacement’ interaction involving physical contact where the assessed animal is butting, hitting, thrusting, striking, pushing or penetrating the receiver with forehead, horns, horn base or any other part of the body with a forceful movement and as a result the receiver gives up its position.

'Penetrating' is defined as an animal pushing itself between two other animals or between an animal and a piece of barn equipment (e.g. the feeding rack, the water trough, a cubicle). If, as a consequence of a displacement, neighbouring animals also leave their feeding places but physical contact as described above is not involved, this reaction is not recorded as displacement.

‘Chasing’ is when the assessed animal makes another animal flee by following fast or running behind it, also by using threats such as jerky head movements. Chasing is only recorded if it follows an interaction with physical contact. If, however, chasing occurs in the context of fighting then it is not counted separately.

‘Fighting’ occurs when two contestants vigorously push their heads (foreheads, horn bases and/or horns) against each other while stamping their feet on the ground and both exerting force against each other. Pushing movements from the side are recorded as long as they are part of the fighting sequence. A new bout starts if the same animal restarts fighting after more than 10 seconds or if the fighting partner changes.

‘Social licking’ is when the assessed animal licks any part of the body (head, neck, torso, legs, tail) of another group mate, except for the anal region or the prepuce. If the assessed animal stops licking for more than 10 seconds and then starts licking the same receiver again, this is recorded as a new bout. If the assessed animal starts to lick another receiver or if there is a role reversal between the two animals, this is also considered as a new bout.

‘Horning’ is head play with physical contact between two animals: the animals rub their foreheads, horn bases or horns against one another’s head or neck without obvious agonistic intention. None of the opponents takes advantage of the situation in order to become a victor. It is considered as a new bout if the same animals start horning after 10 seconds or more or if the horning partner changes.

‘Mounting’ is when the animal lifts itself up on its hind legs and jumps with its forelegs onto a group mate either from behind, the side or front. Displacement or chasing-up following mounting are recorded as separate events.

'Tongue rolling' is when an animal repeatedly twists, twirls or swings its tongue in a stereotypic way inside or outside the open mouth, or stretches out the tongue for longer than 10 seconds.

'Panting' is when the animal shows an abnormal and superficial breathing with a high frequency and sometimes with an open mouth. Usually this behaviour is performed by animals in a condition of heat stress.

'Intention to lie down' is when the animal lowers its head and sniffs or paws the ground, repeatedly bending the carpal joint or shifts weight or swings the head. Except for repeated bending carpal joint or weight shifting, at least two of the previous criteria and no lying down afterwards have to be observed in order to record one episode of 'intention to lie down'. A new lying down intention starts when no criteria are shown for at least 10 seconds or when the same animal changes its position.

How are these behaviours interpreted?

Head butts, displacements, chasing, chasing up and fighting are recorded as agonistic interactions. Social licking and horning as 'cohesive' (considered positive) ones.

Individual animal clinical scoring (6)

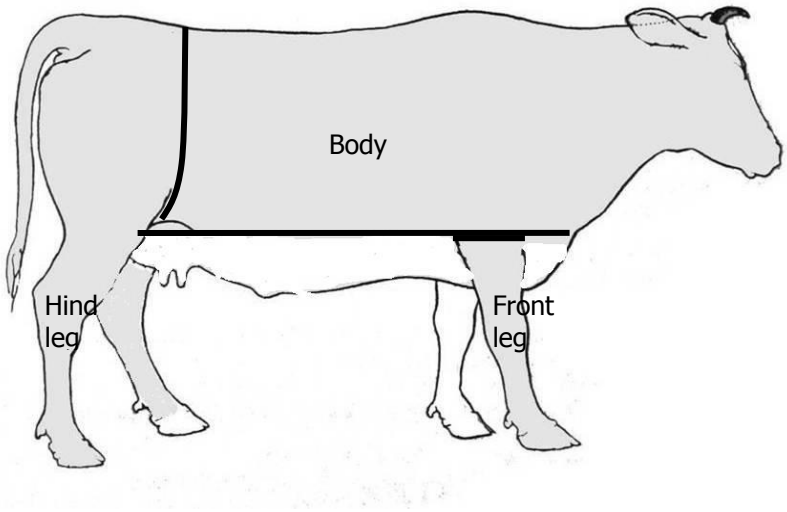
'Integument alterations' is when, in order to assess skin alterations, one side of the animal is examined from a distance not exceeding 2 m, excluding the lower part of the belly and the inner side of the front legs, but including the inner side of the hind leg facing the observer. Generally, the side which is seen 'first' when approaching the animal, is chosen with the goal of achieving an overall random 'side' selection.

The body is divided into 3 regions as shown in Figure A3.

'Skin alterations' of a minimum diameter of 2 cm at the outer margin were counted and distinguished in two categories: 'hairless patches' - areas with hair loss or extensively thinned hair, skin not damaged, hyperkeratosis possible; and 'lesion/swelling' - areas with damaged skin either in the form of a scab or a wound, dermatitis or overt swellings.

All lesions belonging to either category and situated in each body region are counted. If more than 20 alterations per category are present, only '21' is noted and this value, the maximum, is also given if the total area affected is at least as large as the area of a hand, even if this area results by the sum of all alterations. If some body regions are so dirty that they cannot be assessed, a 'D' is noted in the recording sheet. However, if the maximum number of alterations ('21') is seen, this is noted.

Figure A3. Body subdivision for the recording of integument alterations.



For the 'cleanliness' evaluation, the observer assesses the same side of each of the animals as that used for the recording of integument alterations, including as much of the underbelly as is visible, but excluding head, neck and legs below the carpal and tarsal joint. An animal is scored as 'dirty' (score 1) if 25% or more of this area is covered with plaques of dirt. As a modification in this project, an animal is further classified as 'dirty' (score 1) when liquid dirt covers more than half of the considered area, even if plaques were not present. For the recording of the percentage of 'thin animals', Body Condition Scoring is used by assessing the animal's loin, tail head and vertebrae from behind and from the side. To be classified as 'too thin' the animal has to show the indicators shown in Table A12, in 3 of the 4 selected regions.

Table A12. Classification of beef bulls as too thin.

Body region	**Classified as too thin**
Tail head	Cavity around tail head
Loin	Visible depression between backbone and hip bones (tuber coxae)
Vertebrae	Ends of transverse processes distinguishable
General	Tail head, hip bones (tuber coxae), spine and ribs visible

'Nasal discharge' is when flow/discharge from the nostrils is present and clearly visible as a transparent or yellowish/green substance, often of thick consistency. 'Hampered respiration' is when the examined animal shows a deep and laboured breathing, with increased sound and where expiration is supported by the abdominal and thoracic muscles, although the breathing rate could be next to normal or only slightly increased. 'Ocular discharge' is when flow/discharge (wet or dry) from the eye is present, at least 3 cm long, clearly visible. 'Diarrhoea' is defined as presence of an area of wet manure, of at least the size of a hand,

observed below the tail head on both sides of the tail. 'Bloated rumen' is presence of a characteristic 'bulge' between the hip bone and the ribs on the left side of the animal. 'Severe clinical lameness' is observed when the animals are in motion and consists of reluctance to bear weight on a particular foot and frequent weight shifting, to minimise the time of weight-bearing on the affected limb. Lame animals which are standing still can rest a foot, rotate a foot, perform repeated weight shifting ('stepping') or stand on the edge of a step; an animal is classified as 'lame' if 2 of these signs are performed. Front and rear legs are scored separately.

Sampling strategy

The number of pens or animals required for a sample changes with the kind of test and observation. In the case of the beef animal assessment, for herd sizes of up to 100 animals, in addition to assessing 50% of the animals in each pen, all animals are assessed in randomly selected pens until the sample size is reached. With herd sizes of more than 100 animals, the number of 'focal' pens is reduced until the sample size is reached. As far as is practicable, a random sample of pens taking into account distribution across the farm is selected to include a proportion of stock weight classes and also the 'hospital' pens. Behavioural observations are carried out in a sample consisting of a maximum of 12 pens. The total observation time is 120 minutes and each pen is observed for 10 minutes. Pens with more than 25 animals are divided into 2 or more 'segments' in order to safeguard reliable recording.

The range of measures proposed for cattle, poultry and pigs in the Welfare Quality® protocols are detailed in Tables A13 – A21 below.

Table A13. Collection of data for fattening cattle on the farm

	Welfare Criteria		Measures
Good feeding	1	Absence of prolonged hunger	Body condition score
	2	Absence of prolonged thirst	Water provision, cleanliness of water points, number of animals using the water points
Good housing	3	Comfort around resting	Time needed to lie down, cleanliness of the animals
	4	Thermal comfort	As yet, no measure is developed
	5	Ease of movement	Pen features according to live weight, access to outdoor loafing area or pasture
Good health	6	Absence of injuries	Lameness, integument alterations
	7	Absence of disease	Coughing, nasal discharge, ocular discharge, hampered respiration, diarrhoea, bloated rumen, mortality
	8	Absence of pain induced by management procedures	Disbudding/dehorning, tail docking, castration
Appropriate behaviour	9	Expression of social behaviours	Agonistic behaviours, cohesive behaviours
	10	Expression of other behaviours	Access to pasture
	11	Good human-animal relationship	Avoidance distance
	12	Positive emotional state	Qualitative behaviour assessment

Table A14. Collection of data for fattening cattle at the slaughterhouse

	Welfare Criteria		Measures
Good feeding	1	Absence of prolonged hunger	Food supply
	2	Absence of prolonged thirst	Water supply
Good housing	3	Comfort around resting	Flooring, bedding
	4	Thermal comfort	This criterion is not applied in this situation
	5	Ease of movement	Slipping, falling, freeze, try to turn, turn around, move backwards
Good health	6	Absence of injuries	Lameness, bruises
	7	Absence of disease	This criterion is not applied in this situation
	8	Absence of pain induced by management procedures	Stunning effectiveness
Appropriate behaviour	9	Expression of social behaviours	This criterion is not applied in this situation
	10	Expression of other behaviours	This criterion is not applied in this situation
	11	Good human–animal relationship	Vocalizations, coercion
	12	Positive emotional state	Struggling, kicking, jumping in stun box, freezing, trying to turn, turning around, moving backwards

Table A15. Collection of data for broiler chicken on farm (measured on farm)

	Welfare Criteria		Measures
Good feeding	1	Absence of prolonged hunger	This criterion is measured at the slaughterhouse
	2	Absence of prolonged thirst	Drinker space
Good housing	3	Comfort around resting	Plumage cleanliness, litter quality, dust sheet test
	4	Thermal comfort	Panting, huddling
	5	Ease of movement	Stocking density
Good health	6	Absence of injuries	Lameness, hock burn, foot pad dermatitis
	7	Absence of disease	On farm mortality, culls on farm
	8	Absence of pain induced by management procedures	This criterion is not applied in this situation
Appropriate behaviour[1]	9	Expression of social behaviours	As yet, no measure is developed for this criterion
	10	Expression of other behaviours	Cover on the range, free range
	11	Good human-animal relationship	Avoidance distance test (ADT)
	12	Positive emotional state	Qualitative behavioural assessment (QBA)

Table A16. Collection of data for broiler chicken on farm (measured at slaughterhouse)

	Welfare Criteria		Measures
Good feeding	1	Absence of prolonged hunger	Emaciation
	2	Absence of prolonged thirst	This criterion is measured at the farm
Good housing	3	Comfort around resting	This criterion is measured at the farm
	4	Thermal comfort	This criterion is measured at the farm
	5	Ease of movement	This criterion is measured at the farm
Good health	6	Absence of injuries	Breast burns, hock burn, foot pad dermatitis
	7	Absence of disease	Ascites, dehydration, septicaemia, hepatitis, pericarditis, abscess
	8	Absence of pain induced by management procedures	This criterion is not applied in this situation
Appropriate behaviour	9	Expression of social behaviours	As yet, no measure is developed
	10	Expression of other behaviours	This criterion is measured at the farm
	11	Good human-animal relationship	This criterion is measured at the farm
	12	Positive emotional state	This criterion is measured at the farm

Table A17. Collection of data for broiler chicken at slaughterhouse

	Welfare Criteria		Measures
Good feeding	1	Absence of prolonged hunger	Feed withdrawal time
	2	Absence of prolonged thirst	Water withdrawal time
Good housing	3	Comfort around resting	As yet, no measure is developed
	4	Thermal comfort	Panting on lorry and/or lairage
	5	Ease of movement	Stocking density in crates
Good health	6	Absence of injuries	Wing damage, bruising
	7	Absence of disease	Dead on arrival (DOA)
	8	Absence of pain induced by management procedures	Pre-stun shock, effectiveness of stunning
Appropriate behaviour	9	Expression of social behaviours	This criterion is not applied in this situation
	10	Expression of other behaviours	This criterion is not applied in this situation
	11	Good human-animal relationship	This criterion is not applied in this situation
	12	Positive emotional state	Flapping on the line

Table A18. Collection of data for sows and piglets on farm

	Welfare criteria		Measures
Good feeding	1	Absence of prolonged hunger	**Sows:** Body condition score, feeding management **Piglets:** Age of weaning
	2	Absence of prolonged thirst	**Sows and piglets:** Water supply
Good housing	3	Comfort around resting	**Sows:** Bursitis, shoulder sores **Sows and Piglets:** Absence of manure on the body
	4	Thermal comfort	**Sows and Piglets:** Panting, huddling
	5	Ease of movement	**Sows:** Space allowance, farrowing crates
Good health	6	Absence of injuries	**Sows and piglets:** Lameness **Sows**: Wounds on body, vulva lesions
	7	Absence of disease	**Sows and piglets:** Mortality, coughing, sneezing, pumping, rectal prolapse, scouring **Sows:** Constipation, metritis, mastitis, uterine prolapse, skin condition, ruptures and hernias, local infections **Piglets**: Neurological disorders, splay leg
	8	Absence of pain induced by management procedures	**Sows:** Nose ringing and tail docking **Piglets:** Castration, tail docking and teeth clipping
Appropriate behaviour	9	Expression of social behaviours	**Sows:** Social behaviour
	10	Expression of other behaviours	**Sows:** Stereotypies, exploratory behaviour
	11	Good human–animal relationship	**Sows:** Fear of humans
	12	Positive emotional state	**Sows and piglets:** Qualitative behaviour assessment (QBA)

Table A19. Collection of data for growing pigs on farm (measured on farm)

	Welfare criteria		Measures
Good feeding	1	Absence of prolonged hunger	Body condition score
	2	Absence of prolonged thirst	Water supply
Good housing	3	Comfort around resting	Bursitis, absence of manure on the body
	4	Thermal comfort	Shivering, panting, huddling
	5	Ease of movement	Space allowance
Good health	6	Absence of injuries	Lameness, wounds on body, tail biting
	7	Absence of disease	Mortality, coughing, sneezing, pumping, twisted snouts, rectal prolapse, scouring, skin condition, ruptures and hernias
	8	Absence of pain induced by management procedures	Castration, tail docking
Appropriate behaviour	9	Expression of social behaviours	Social behaviour
	10	Expression of other behaviours	Exploratory behaviour
	11	Good human–animal relationship	Fear of humans
	12	Positive emotional state	Qualitative behaviour assessment (QBA)

Table A20. Collection of data for growing pigs on farm (measured at the slaughterhouse)

	Welfare criteria		Measures
Good feeding	1	Absence of prolonged hunger	This criterion is recorded at the farm
	2	Absence of prolonged thirst	This criterion is recorded at the farm
Good housing	3	Comfort around resting	This criterion is recorded at the farm
	4	Thermal comfort	This criterion is recorded at the farm
	5	Ease of movement	This criterion is recorded at the farm
Good health	6	Absence of injuries	This criterion is recorded at the farm
	7	Absence of disease	Pneumonia, pleurisy, pericarditis, white spots in the liver
	8	Absence of pain induced by management procedures	This criterion is recorded at the farm
Appropriate behaviour	9	Expression of social behaviours	This criterion is recorded at the farm
	10	Expression of other behaviours	This criterion is recorded at the farm
	11	Good human–animal relationship	This criterion is recorded at the farm
	12	Positive emotional state	This criterion is recorded at the farm

Table A21. Collection of data for finishing pigs at slaughterhouse

	Welfare criteria		Measures
Good feeding	1	Absence of prolonged hunger	Food provision
	2	Absence of prolonged thirst	Water supply
Good housing	3	Comfort around resting	Flooring, bedding
	4	Thermal comfort	Shivering, panting, huddling
	5	Ease of movement	Slipping, falling, stocking density of lorries, stocking density of lairage pens
Good health	6	Absence of injuries	Lameness, wounds on body
	7	Absence of disease	Sick animals, dead animals
	8	Absence of pain induced by management procedures	Stunning effectiveness
Appropriate behaviour	9	Expression of social behaviours	This criterion is not applied in this situation
	10	Expression of other behaviours	This criterion is not applied in this situation
	11	Good human–animal relationship	High pitched vocalizations
	12	Positive emotional state	Reluctance to move, turning back

A5. How could an assessment tool, such as the Welfare Quality tool, which includes ABMs, be used to raise the overall level of animal welfare?

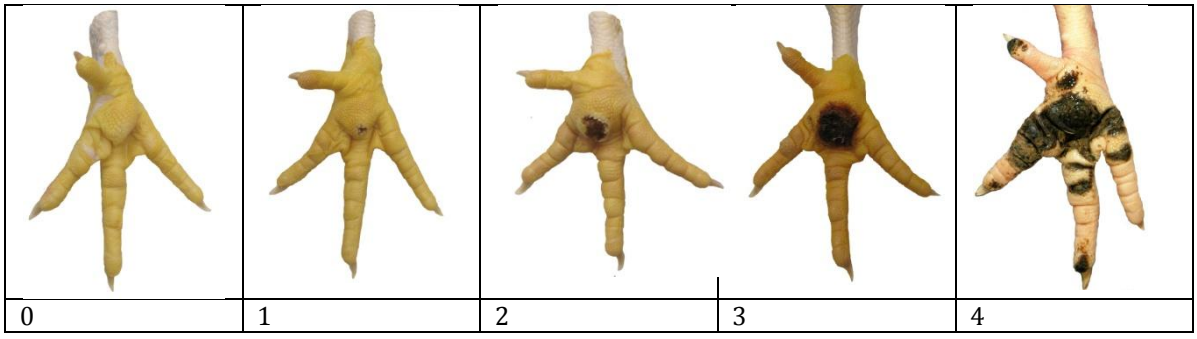
There are a number of 'levels' to application of an ABM. First, a standardised description is needed, then a description of the practical application of the method and any linked RBMs is required. Then, finally, the ways in which the information that is produced from the use of the ABMs and RBMs can be used to promote and support management decisions and practices to create improvements in welfare:

Step 1: Measure (ABMs and RBMs) leading to
Step 2: Analyse risk factors leading to
Step 3: Inform (producer, purchaser) leading to
Step 4: Support management decisions to create improvements in welfare.

Now, let us give an example, using the measure: foot pad dermatitis in broiler chicken. Description: foot pad dermatitis (or pododermatitis) is a contact dermatitis found on the skin of the foot, most commonly on the central pad, but sometimes also on the toes. The skin is turned dark by contact with litter and deep skin lesions can result. The scoring scale allows an assessment of the severity of these lesions (Figure A4). The condition is comparatively common, but the incidence is variable with some flocks showing much higher levels than others.

In this example of foot pad dermatitis, at least 100 broiler chicken per flock should be assessed: 10 birds taken from 10 areas of the house including 2 areas located near to drinkers, 2 located areas near to feeders, 3 areas located near a wall, 3 areas located away from drinkers and feeders (resting area). At the slaughter plant, it is possible to score a large number of feet as the birds pass on the slaughter line. If the line speed is 120 birds per minute, then observing for 10 minutes will enable scoring of 1200 birds using the same scoring scale as used on farm.

Figure A4. Scoring scale for pododermatitis (foot pad disease) in broiler chicken.



0	1	2	3	4

Step 1 - Training assessors to carry out the following measures:

- training in scoring for pododermatitis is by means of on-farm visits and examination of photographs;
- the inspectors will require to be assessed during the training course until they develop a uniform scoring;
- assessors will also be asked to periodically carry out a validation or reference audit to check that they continue to score in a repeatable way.

Assessors should pen approximately 25 birds in the house. Birds should not be herded into and pen and the pen should be placed quietly around a group of birds with minimal disturbance. If birds show any signs of distress or overheating they should be released and a new group penned. Then they should pick up individual birds carefully and score them and place them outside the pen. Then they should create a simple tabulation of the numbers of birds scored in each foot pad category.

As well as recording the ABM (foot pad scores), a farm questionnaire or standard inspection report with the assistance of a representative of the farm will need to be completed which provides a description of the farm, house and flock (the RBMs) which gives the background and risk factors required to provide information to help the farmer resolve his foot pad dermatitis issues. This questionnaire should include: broiler breeder information including genotype/strain, broiler breeder history and age; hatchery information including distance/time transported and hatchery vaccination programme for chicks; general information including number and weight of chicks placed, sex, time of year, age at assessment and slaughter; specific husbandry practices including stocking density and thinning practice, brooding conditions, nutritional profile, vitamin/mineral levels, litter substrate, feeder and drinker design/type, lighting programme, age of house, construction details, target ventilation profile, diseases and medication history, vaccination programme and water source; performance information including growth profile, weekly mortality pattern, weekly leg cull pattern, other culls weekly pattern; and background information about the site/company including size of houses, number of birds on site and biosecurity measures. Nevertheless, it should be pointed out that many of these RBM questions form the basis of existing farm assurance inspection.

Step 2 - Analyse risk factors:

The farm or company (or the assurance or advisory service) can (in this specific example) analyse the ABM and RBM information and:

- Find out the prevalence and severity of foot pad dermatitis in the flocks within a company. Include in this an assessment of the economic impact of small birds, moribund birds and culled birds which result from lameness. In general, significant improvements in profitability can be made if floor pad problems are tackled within a company, as well as an improvement in overall bird welfare - inspection bodies in some countries are now beginning to focus on foot pad health as a marker for company welfare performance.
- Make comparisons between ‘good’ and ‘poor’ farms (with respect to foot pads) within a company to help identify management, house environment, feeding, medication, stockmanship and genotype factors which differ between these farms.
- Investigate the use of water in different farms - farms with increased water use per bird may have systematic problems with leaking drinkers. Small amounts of water leaked chronically into the litter can severely damage the management of litter and foot pad health.
- Carry out an investigation of the bacteriological pathologies linked with foot pad dermatitis and identify whether these bacteria can originate in the hatchery, the transportation or by lapses in farm biosecurity.

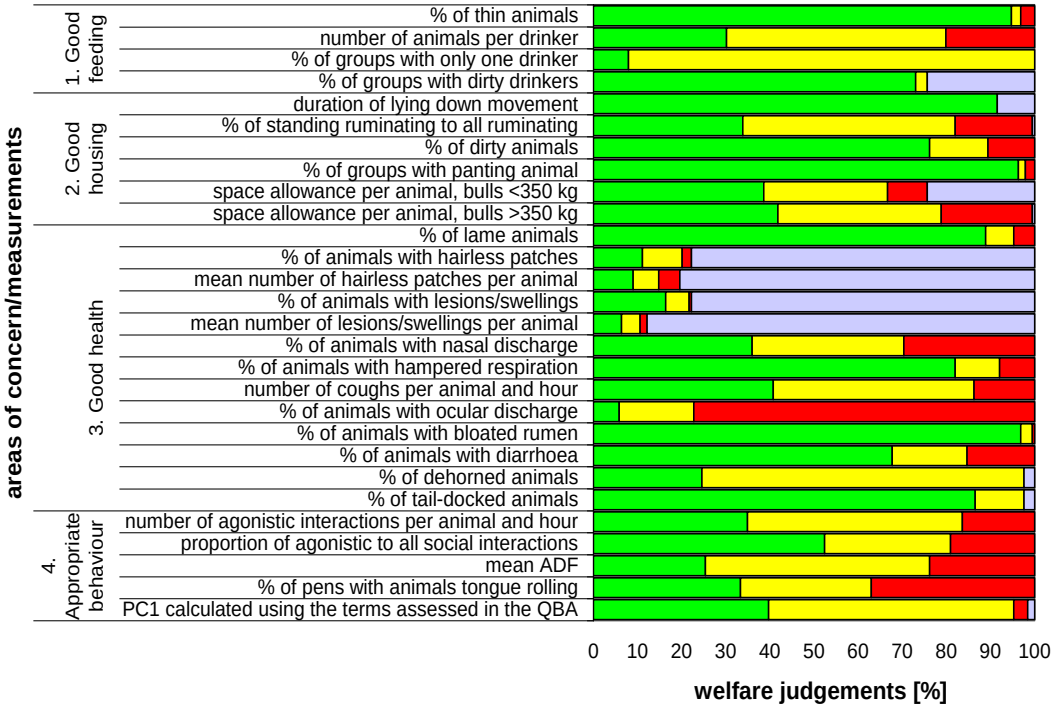
Step 3 - Inform and Step 4 - Support management decisions to create improvements in welfare:

The farmer can be informed about the extent of foot pad dermatitis, and, with time, and after analysis, a pattern of risk factors may emerge which allow him to make decisions which can reduce this. Factors which, in real farm experience, have been shown to be risk factors for lameness include: growth rate, the age of the birds at slaughter, the use of whole cereals in the diet, the type of feed, the quality of biosecurity measures, litter condition and the genotype of the birds. Also, the gender of the birds, levels of feed restriction, the lighting pattern and light intensity, bird activity levels and the stocking density have, on farms, been altered and manipulated to control levels of foot pad dermatitis in the chickens.

A6. Welfare Quality Example: scoring of the welfare state of beef farms

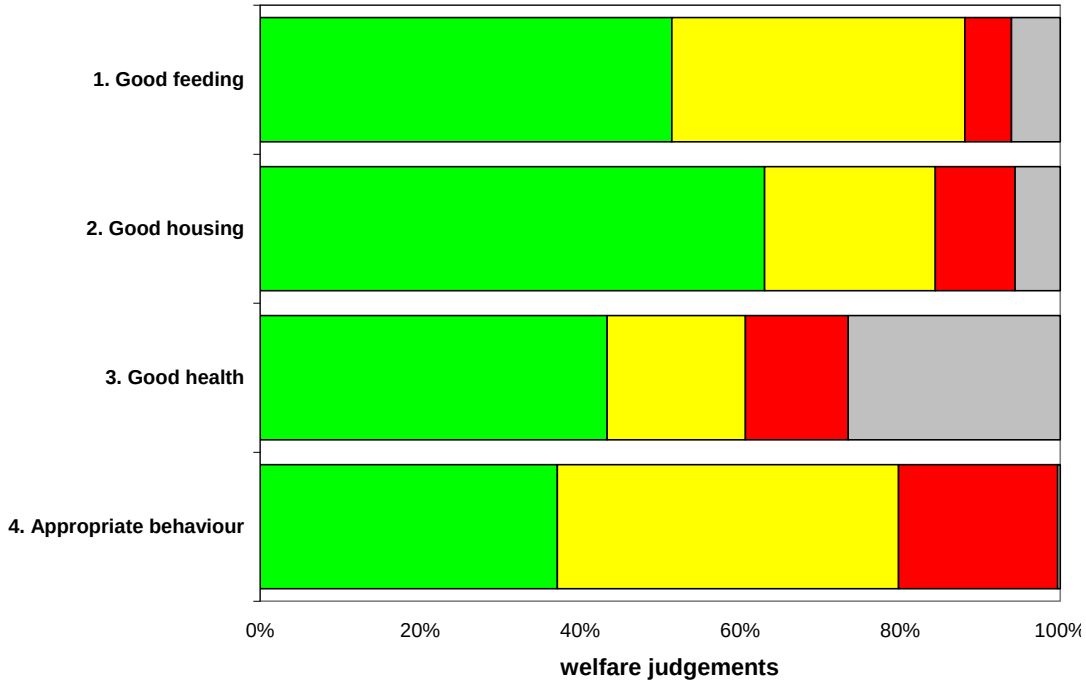
To illustrate the range of 'real assessment data' that is seen when the Welfare Quality system is used on farm, an overview over the welfare states of the animals on 189 different beef farms from a Welfare Quality assessment is presented in Figures A5, A6 and A7. The origin of these farms is not revealed - they are presented simply to demonstrate the spread of values seen if the scheme is applied.

Figure A5. Welfare assessment results from 189 beef farms.



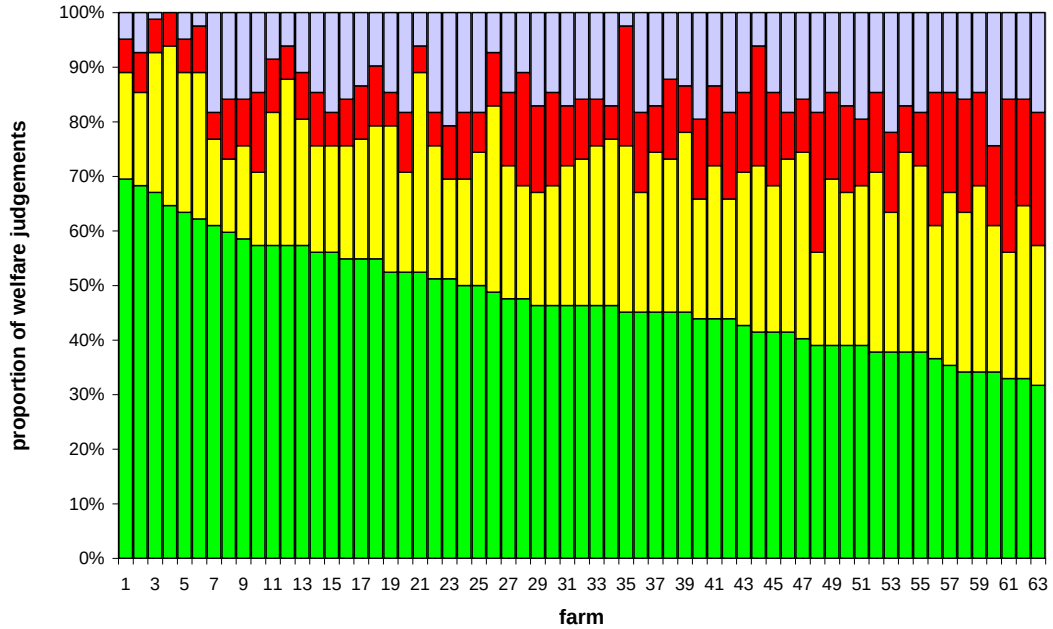
For each measure (green: 'ok'; yellow: 'improvement desirable'; red: 'improvement necessary'; grey: 'missing value'). % of dehorned and tail-docked animals were only recorded twice (n=126); for the latter measures as well as % of groups with only one drinker and % of groups with dirty drinkers only the categories 'ok' and 'improvement desirable' were used.

Figure A6. Proportion of welfare assessments for 189 farms as classified in the four major categories.



1. Good feeding; 2. Good housing; 3. Good health; 4. Appropriate behavior; for all farms and all assessments for each area of concern (n=189; green: 'ok', yellow: 'improvement desirable', red: 'improvement necessary', grey: 'missing value').

Figure A7. Ranked proportion of welfare categories for 63 beef farms - listing overall assessments and measures (green: ok', yellow: 'improvement desirable', red: 'improvement necessary', grey: 'missing value).



A7. Potential veterinary use of animal based assessment methods

Veterinarians in Europe play important roles in relation to animal welfare issues. Vets who work in practice interact on a day-to-day basis with their farming clients, diagnosing and treating animal disease and advising farmers on disease prevention and reduction, animal behaviour, welfare issues, breeding and nutrition (Butterworth *et al.*, 2008). Some vets are responsible for both animal and public health considerations at slaughterhouses, and help to ensure that animals are treated well at the time of slaughter. Others carry out surveillance for both animal and zoonotic disease in farm and companion animals, and during transport and at international borders. As well as assessment of health and welfare, vets can assist farmers to implement strategies to improve welfare through improvements in the quality of stockmanship, improvements in housing and husbandry, and through genetic selection.

Despite a large amount of co-operation between veterinary organisations and government agencies, there is comparatively little 'benchmarking' information shared between countries when it comes to animal welfare parameters. In all countries of the EU, the state, usually through a state veterinary service carries out some visits to farms to ensure that they can deliver compliance with animal care and welfare legislation. It is possible that demonstration of adequate (or even high) welfare standards would allow targeted farm inspection, and the state inspection load could be optimised using information provided by inspection schemes which included a harmonised welfare assessment. The potential for streamlining of animal health visits could be a genuine benefit to farmers and to the state.

Let us consider an example. A dairy farmer has a problem with lameness in his dairy cows. A structured assessment will tell him how many lame cows he actually has when referenced to some standard definitions. Once he has a tool to gauge whether the level of lameness is increasing (or decreasing) then this can be a barometer against which to judge which practical steps he may make to reduce the problem. ABMs may help him to identify practical ways of trying to reduce lameness. For example, information on the type of floor and the farmer's hoof care strategy could be used to help advise on remedial solutions to improve the problem which can be both an economic cost and also a cost to the animals in terms of disability or discomfort. Targeted improvement may be able to help both the farmer and the animal. To be viable, remedial strategies must satisfy both welfare and economic requirements, and they must be practicable.

Alongside this advisory information, the vet is also able to tell the farmer how his animals are doing with regards to early calf mortality, to respiratory disease, to udder lesions and with respect to the feeding of his young cattle. When this information is linked to the economic information that the farmer is likely to share with his vet, then the vet may have a powerful tool to help and support his farming clients and to promote best management in animal health and welfare. This productive relationship between the vet and the farmer relies on several things - trust, information and economic sense. Trust comes from the professional interaction between vet and client, information is part of the armoury of skills that makes the

vet a valuable part of the farming system, knowledge of what is common and uncommon, what disease looks like, the specialist techniques of surgery and medicine and knowledge of how his clients farms, and economic, disease and welfare performance relates to other similar farms. Farm veterinary services are not alone in benefiting from benchmarked indices on which to base management decisions, the use of Key Performance Indicators is a cornerstone of good business management in many commercial practices. The feedback of information to the farmer, often with veterinary involvement, and his/her uptake of recommendations and remedial measures, may represent a potential direct advantage of this approach. Veterinarians who work directly with their farm clients, those involved in assurance schemes, those carrying out disease surveillance and import and export work may find that a series of harmonised welfare assessment tools are of value to their business. These tools may be of use for a number of purposes:

- to allow inspection and scoring of farms ;
- to inform consumers about the welfare status of the animals from which they buy products ;
- to provide advice and support for the farmer based on the data collected on the farm; and
- additionally, technical specifications may be used by veterinary inspection agencies in their statutory assessment of farms.

A8. The importance of assurance schemes and their inspectors in the evolving welfare assessment picture

Assurance schemes in a number of countries have the potential to influence and improve farm animal welfare through their programmes of inspection to standards. Higher welfare animal products, sometimes carrying the label of an assurance scheme, are generally sold to consumers as 'quality items' or as 'best' or 'finest', thereby generating a higher price (Buller & Roe, 2009).

Critical to the audit process is the relationship between the farmer and the assessor. Assessors are 'not the police' and, depending on the nature of the assurance scheme, their relationship to farmers usually remains at the level of professional detachment, in which advice may not be given (schemes which operate under a harmonized certification standard EN45011 may not offer advice associated directly with the assessment). However, some schemes, most notably the Soil Association in the UK have set up a distinct advisory arm. Additionally, in a number of countries, it appears common for a 'farm advisor' to offer the producer support and advice in preparation for an assessment audit (this person is distinct from the assessor).

However, the practical application of ABMs within existing assurance schemes could be problematic in a number of ways, for example:

- are existing schemes able to assess ABMs in a repeatable and reliable way within the timescale of an routine assessment?

- who will carry the cost of assessing ABMs?

- how will assessment of ABMs work in terms of periodicity and seasonality of assessment i.e. will the seasonal changes in production systems make interpretation of the findings difficult?

- can reduction of ABMs to a single farm-based score actually work?

- ABMs conventionally assess negative findings (lameness, skin lesions, hunger), can they also be used to convey 'positive' information to consumers?

Alongside these considerations, the incorporation of ABMs into an existing assessment framework which mostly uses RBMs will include practical questions such as:

- to 'integrate ABMs into existing standard farm assessments' or 'to keep them separate"?
- can use be made of slaughterhouse derived ABMs when creating farm assessment reports? and
- can vets assess ABMs?

Appendix B. Assessment of the total potential consumer spend on animal welfare

B1. An estimate of the market size of brands and products with animal welfare attributes

Over the last 10 years or so, a major feature in the retail trade has been the growth in so-called 'ethical consumerism' or 'ethical shopping', as the Institute of Grocery Distribution (2010a) has described it. Definitive figures on the annual spend on such products do not exist and some of the data that has been published needs to be treated with caution due to the nature of the publisher e.g. individual retailers and pressure groups who may not be impartial. However, despite this caveat, we will here attempt to present estimates of the market size of brands and products with animal welfare attributes.

According to the Co-operative Bank (2009 & 2010), spending on ethical food and drink rose by almost three times from £1.9 billion in 1999 to over £6.5 billion in 2009. To set this in context, it should be noted that 'fair trade', 'organic', 'free range' and 'dolphin friendly' products were available before 1999 but with limited availability. Furthermore, since then, various new certification schemes covering such products have been introduced and widely established, such as the RSPCA's animal welfare-based Freedom Food brand and the Marine Stewardship Council's eco label for fish. Yet, according to the Institute of Grocery Distribution (2010b), 31% of the population expected to purchase more products with animal welfare attributes in the year ahead. Table B1 presents data, for the last decade, on consumer spending in the UK on brands and products that could be described, in the broadest sense, as possessing animal welfare attributes.

Table B1. Consumer spending on brands and products with animal welfare attributes, UK, 1999 and 2009 (£ million)

	1999	2009
Organic food	390	1704
Free range eggs	173	447
Free range poultry meat	37	174
Freedom Food products	-	122
Dolphin friendly tuna	189	281

Source: Derived from the Co-operative Bank (2009 & 2010)

Organic food, as seen from Table B1, not only had the highest spend in both years, but also the highest proportionate growth between 1999 and 2009. However, as this definition will include organic fruit, vegetables and cereals, only a part is likely to possess animal welfare attributes – organic meat, dairy products and egg production because of the certification schemes' stipulation of free access of the animals to open range.

No official figures exist for what the value of such spending is. However, Lampkin *et al.* (2008) state that, in 2007/8, expenditure on organic meat, dairy products and fish was £330 million. They did not estimate the value of organic eggs bought. So, perhaps, the value of organic food with animal welfare attributes could be of the order of £500 million per year. Whilst the purchase of dolphin friendly tuna is indicative of public attitudes towards animal welfare, and the spend in 2009 of £281 million is substantial, it might not be correct to include it in our computation of consumer spend on animal welfare brands and products.

Of the brands and products listed in Table B1 that can be described as 'genuinely' possessing animal welfare attributes, free range eggs, free range poultry meat and Freedom Food products account for spend of £743 million in 2009, up by some 350% from 1999.

A further indicator of consumers' interest in food quality and safety, and also their concern over production methods, is the rise in the numbers of farms taking part in the main baseline UK farm assurance schemes (Assured Food Standards) which also carry the 'Red Tractor' logo. Such schemes relating to animal production usually have a 'minimum' farm animal welfare stipulation and, according to Assured Food Standards (2009), the value of all retail food sales produced under these schemes rose from £4.5 billion in 2003/04 to £8.5 billion in 2008/09. Defra (2008) state that the poultry and dairy livestock sectors had the highest proportion of assured production at 95% of total animals with sheep producers the least represented at 65%.

Of the farm assurance schemes that exist, the RSPCA Freedom Food Scheme is specifically aimed at giving consumers confidence in the animal welfare provenance of the food they buy. As shown in Table B2, the rise in the number of farm animals kept under this Scheme has been spectacular; from 22,929,000 head in 2004 to 372,901,000 in 2009. Of those animals shown in the table, the largest proportionate rise in this period was in the number of broilers under the Scheme.

Table B2. The total number of animals farmed in the UK under the RSPCA Freedom Food Scheme ('000 head)

	All animals	Laying hens	Broilers	Pigs
2004	22929	9891	10069	1525
2005	43487	11034	22687	1564
2006	51081	11623	25144	1477
2007	153760	12610	44353	1373
2008	356581	14371	55097	1574
2009	372901	14786	56919	2021

Source: RSPCA's Freedom Food Ltd, personal communication.

Defra (2008) give important market penetration figures for Freedom Food sales for England for 2007. These show that laying hens had the highest penetration at 52.1%, followed by ducks with 24.8% and pigs at 14.5%; less than 1% of beef and sheep meat were under the

Freedom Food brand. When these figures are considered, and taken together with those in Table B2 from the RSPCA themselves, this casts doubts on the consumer spend figures presented above in Table B1 for Freedom Food products at £122 million which seems to have to be an under estimate.

B2. Review of charitable giving associated with animal welfare

By repute the British public is a generous giver of money to charitable causes; in 2008/09, according to the Charities Aid Foundation/National Council for Voluntary Organisations (2009), they donated £10.6 billion, an amount which was up from £9.1 billion in 2004/05. However, charities receive their funds from a variety of sources not just donations, legacies and gifts from the public. Other sources include membership fees, charitable activities, trusts and grants from local and national government. When these are added together a total income sum will result but, perhaps surprisingly, no such sum is available at present.

Nevertheless, various 'spend' figures for charities have been estimated, the most accurate being that provided by Caritas Data (2009) for the top 3,000 charities in the UK of £33 billion. But, how much of this spend is on animal welfare activities or causes and, of this, what proportion is on farmed animal welfare?

The Charity Commission (2010), a NDPB, listed the annual income ranges of all the registered animal charities. A summary of this data is shown in Table B3.

Table B3. The number of UK registered animal charities and their annual income ranges.

Income range	No. of charities
£0 – 10,000	64
£10,001 – 100,000	113
£100,001 – 1,000,000	89
£1,000,001 and over	19
	285

Source: Derived from the Charity Commission (2010).

It can be seen that, whilst 177 of the total 285 registered animal charities had annual incomes of up to £100,000, there were 19 such charities with incomes of over £1,000,000. To calculate the total annual spend from this data would require several heroic assumptions, especially trying to calculate the proportion of spending that was for animal welfare itself.

Luckily, for our present purposes, Caritas Data (2009) calculated that the annual spend of animal welfare charities in 2009 was £175 million in the UK, some 0.5% of the total spend of the top 3000 charities. (Sectors receiving the most support were health and medical charities and those engaging in international development.) The Charities Aid Foundation/National Council for Voluntary Organisations (2007) found that animal welfare causes receive a higher

proportion of giving by the public than they do of trust and foundation grants, and that 80% of the giving to animal welfare charities in 2006/07 came from low-level donors and that women are more likely to support these causes than men.

Table B4. The income of the main animal welfare charities operating in the UK, 2007/08 (£ million)

Charity	Income
RSPCA	114.1
WSPA	23.4
Donkey Sanctuary	20.5
PETA	19.0
Soil Association	15.0
The Brooke	11.1
CIWF	4.3

Source: Ecos Consultancy (2009).

The biggest UK charity concerned with farmed animal welfare is the RSPCA with an annual income in 2007/08 of £114.1 million. As shown in Table B4, this is five times the next highest annual income of an animal welfare charity – that of the WSPA with £23.4 million and over six times that of the next one, the Donkey Sanctuary, with £20.5 million. The Brooke and the Donkey Sanctuary focus on horses and donkeys so the Ecos Consultancy (2009) consider them to 'be on the boundary between farmed and companion animals'.

What the spend on farmed animal welfare amounts to is difficult to say, although the Ecos Consultancy (2009) suggest that it is 'a small fraction'. Furthermore, they state that 11% of the RSPCA spend goes on this each year – some £12.6 million in 2007/08. If we apply this same proportion to the income figure of the WSPA, the Soil Association and CIWF, we get a further sum of £4.7 million, making a possible total spend on farmed animal welfare of the main UK charities each year of some £17 million.

B3. Estimating the total consumer spend on animal welfare

In Section B1, we presented estimates of the market size of brands and products with (farm) animal welfare attributes. Making such estimates often involved heroic assumptions as did the estimates presented in Section B2 reviewing charitable giving associated with animal welfare. Both sets of estimates are summarised in Table B5 below against a judgement on the reliability of the estimates.

It can be seen that the total of almost £1,699 million estimated for 2009 is made up of the £175 million charitable spend associated with animal welfare and, £1,524 million for 5 broad categories of brands and products with animal welfare attributes. It should also be pointed out that, with the exception of the amount spent on dolphin-friendly tuna, the majority of

this latter category went on farmed animal welfare, whereas probably only around 10% of the charitable spend associated with animal welfare was on farmed animal welfare.

Table B5. An assessment of the total potential consumer spend on animal welfare, UK, 2009 (£ million)

	Spend	Reliability of estimate[1]
Brands and products with animal welfare attributes		
Organic meat, dairy products and eggs	500	?
Free range eggs	447	**
Free range poultry meat	174	*
Freedom Food products	122	?
Dolphin friendly tuna	281	**
Charitable spend[2] associated with animal welfare	175	*
Total	1699	*/?

[1] ** for very reliable; * for reliable; and ? for a rough estimate.

[2] This is spend not income; it can be seen in Table B4 that the combined income of the 7 main UK animal welfare charities is £207.4 million but this includes grants from the public sector as well as donations from companies.