European Food Safety Authority

EFSA Journal 2012;10(6):2767

SCIENTIFIC OPINION

Statement on the use of animal-based measures to assess the welfare of animals[1]

EFSA Panel on Animal Health and Welfare (AHAW)[2, 3]

European Food Safety Authority (EFSA), Parma, Italy

ABSTRACT

The overall aim of the work outlined in this statement is to help establish a common framework for future scientific opinions dealing with the use of animal-based measures to assess the welfare of animals. The statement is mainly intended to support the work of EFSA, and a list of considerations for experts working on these future scientific opinions is presented. EFSA has already published a scientific opinion on dairy cattle and another on pigs related to the topic. This statement clarifies some common issues on terminology and integration of concepts, and presents some essential characteristics of animal-based measures to ensure that they are 'fit for purpose'. It highlights that more information is needed about the direction and strength of the various links between input factors and the animal-based measures (welfare indicators) that are used to assess their consequences. The statement highlights the importance of the systematic collection of standardised field data on animal-based measures and subsequent availability in well-defined databases. Targeted analysis of such data will help when selecting the most appropriate measure, or combination of measures, from the 'toolbox' of many potential measures, according to the specific purpose of the welfare assessment, as well as contribute to better assessing their validity and robustness. This will support the move towards quantitative risk assessment of animal welfare.

KEY WORDS

Animal welfare, animal-based measures, welfare outcome indicators, risk assessment, welfare assessment

---

[1] On request from EFSA, Question No EFSA-Q-2012-00275, adopted on 23 May 2012.

[2] Panel members: Anette Bøtner, Donald Broom, Marcus G. Doherr, Mariano Domingo, Jörg Hartung, Linda Keeling, Frank Koenen, Simon More, David Morton, Pascal Oltenacu, Fulvio Salati, Mo Salman, Moez Sanaa, James M. Sharp, Jan A. Stegeman, Endre Szücs, Hans-H. Thulke, Philippe Vannier, John Webster and Martin Wierup. Correspondence: ahaw@efsa.europa.

[3] Acknowledgement: The Panel wishes to thank EFSA staff: Eleonora Bastino, Franck Berthe, Chiara Fabris, Maria Ferrara and Tomasz Grudnik for the support provided to this scientific opinion.

Suggested citation: EFSA Panel on Animal Health and Welfare (AHAW); Statement on the use of animal-based measures to assess the welfare of animals. EFSA Journal 2012;10(6):2767. [29 pp.] doi:10.2903/j.efsa.2012.2767. Available online: www.efsa.europa.eu/efsajournal

SUMMARY

EFSA has been asked to produce a series of scientific opinions on the use of animal-based measures to assess the welfare of farm animals. The starting points for the first of these opinions have been the previous EFSA opinions on the welfare of dairy cattle, pigs and poultry and the protocols developed in the EU-funded Welfare Quality® project and other similar scientific research. Some issues arose in the process of combining the risk assessment approach and the welfare assessment approach. This statement summarises some of these issues and so contributes towards establishing a common framework for future EFSA scientific opinions dealing with the use of animal-based measures in the assessment of welfare, both as welfare outcome indicators and for animal welfare research. This statement also discusses the potential for access to systematically collected, animal-based measures for EFSA's future work in animal welfare risk assessment.

The scientific opinions published so far in the series have proposed lists of potential animal-based measures that could be linked to hazards in the animal's environment and used to determine the extent to which recommendations in earlier EFSA opinions have been fulfilled. However, these lists of animal-based measures are long and it is not necessary to recruit all measures on every occasion when the welfare of an animal is to be assessed. Instead it is proposed that they are considered as a form of "toolbox", from which to select the range of animal-based measures necessary to address the specific objectives of the assessment for that particular species and category of animal at that time. That is to say, the measures chosen should be "fit for purpose". Which measure is the most appropriate will depend on a number of different things, e.g. the purpose of the assessment, the skills of the person collecting the measure, the conditions under which it is to be gathered, the time available to collect it and financial constraints.

In many respects the issues relevant to using animal-based measures to assess welfare are comparable to those relevant to using diagnostic tests of disease to assess health. The second part of this statement focuses on the criteria, such as validity, sensitivity, specificity and robustness, that should be used to select appropriate animal-based measures. The systematic collection of field data on animal-based measures and subsequent storage in well-defined databases can in the future assist in better assessing their potential and their limitations as animal welfare outcome indicators. It could also contribute significantly to visualising the interactions between animal-based measures and the links between animal-based measures and the factors in the environment of the animal that affect them, thereby helping future decisions about optimal combinations of measures. For example, it may be possible to further reduce the number of selected measures needed for an overall welfare assessment if there is overlap in the information contained. Achieving this implies identifying and exploring not only these links and interactions but also their strength and their predictive capacity.

The work carried out so far while developing the scientific opinions on animal-based measures has shown that because of the complexity of the system and the heterogeneity of unlinked data sources it is currently difficult to combine information, originally collected for different aims, into a complete picture. EFSA's opportunities to move towards quantitative risk assessment are currently limited by insufficient systematically collected field data, at the animal, herd and farm level, captured in a centralised database, from which to explore interactions between hazards and animal-based measures of consequences. Besides providing important data for EFSA, the systematic recording of a few standardised valid and robust animal-based measures could be part of an animal welfare surveillance scheme, which would have other benefits. For example, benchmarking of important animal-based measures on a large scale would give quicker feedback to policy makers on the effectiveness of legislation or other initiatives to improve animal welfare as well as be a management tool for the animal industries.

Given that such standardised field data are not yet available, the following steps are proposed to be included in future EFSA opinions on the use of animal-based measures to assess welfare: (1) identification of the animal-based measures to be placed in the toolbox for each animal species; (2) critical examination of the essential characteristics of the animal-based measures in the toolbox; (3) selection of a shortlist of animal-based measures for the purposes of the welfare assessment and

harmonised collection of these, (4) promotion of the establishment of a database for future quantitative risk assessment of animal welfare and identification of a set of resource, management and animal-based measures that should be recorded for this purpose; and (5) emphasis of the use of good operating procedures and reporting standards.

TABLE OF CONTENTS

BACKGROUND

The Community Action Plan on the Protection and Welfare of Animals (2006–2010) identified as two main areas of action, "upgrading existing minimum standards for animal protection and welfare. …" and "introducing standardised animal welfare indicators in order to class the hierarchy of welfare standards applied …".

The EU-funded Welfare Quality® project had, as main outcomes, a science-based methodology for assessing animal welfare and a standardised way of integrating this information to assign cattle, pig, or poultry farms to one of four animal welfare categories (from poor to excellent).

The recently adopted EU Strategy for the Protection and Welfare of Animals (2012–2015) highlights that the possibility of using scientifically validated outcome-based indicators complementing prescriptive requirements in EU legislation will be considered when necessary.

The use of animal-based measures, however, to assess animal welfare-outcomes for legislative purposes is relatively new, although diverse research projects now focus on these, and such measures are also considered in various assessment schemes. Animal-based measures aim to measure the actual welfare of the animal and thus include the effects of different input factors.

EFSA has been requested to produce different scientific opinions on the use of animal-based measures to assess the welfare of farm animals. The first request concerned the use of animal-based measures to assess the welfare of dairy cows (M-2010-0263). This was followed by requests on the assessment of pig welfare (M-2011-0131) and boiler welfare (M-2011-0227). The scientific opinions should give an independent view of the use of animal-based measures in those species, through the joint evaluation of the conclusions and recommendations of the previous EFSA scientific opinions and the assessment protocols suggested by the Welfare Quality® project and other scientific literature.

During the development of the first of these scientific opinions on animal-based measures for dairy cows, the AHAW Panel identified the need to address general issues related to the use of animal-based measures and the development of tools to assess animal welfare.

The statement presents general concepts and principles related to the use of animal-based measures and the development of tools for welfare assessment and monitoring, establishing a common framework for specific and detailed opinions on welfare assessment for a wide range of species. The repetition of a generic section in future scientific opinions will thus be avoided.

TERMS OF REFERENCE

The statement will provide a general overview on the use of animal-based measures to assess welfare of farm animals and a general concept for the development of tools to monitor animal welfare.

ANALYSIS

1. INTRODUCTION

EFSA has been requested to produce different scientific opinions on the use of animal-based measures to assess the welfare of farmed animals. The first request on the use of animal-based measures to assess the welfare of dairy cows was followed by similar requests on the assessment of pig welfare and broiler welfare. The starting points were the previous EFSA opinions on these species and the protocols developed in the EU-funded Welfare Quality® project. This statement summarises some of the general issues that arose during the development of the resulting three scientific opinions and puts them into the broader context of EFSA's future work in the assessment of animal welfare risks and benefits.

The first section of this statement outlines the risk assessment approach of EFSA, where the focus is on identifying input factors, and the welfare assessment approach, where the focus is on identifying welfare outcomes. The concepts of animal welfare used in risk assessment and welfare assessment overlap considerably. However, the two approaches had not been integrated previously, at least not in the way that was necessary to proceed with the scientific opinions, and some of the conceptual problems that arose in the process of combining them are expanded upon.

The resulting scientific opinions propose lists of potential animal-based measures that could be used to assess the welfare of, for example, dairy cattle and pigs (EFSA, 2012a, b). But it was recognised in all these opinions that much work is still needed to standardise the measures and the methods used to record them. Thus, the second part of this statement focuses on the criteria that should be used to select the most appropriate animal-based measures from a "toolbox" of potential measures. The final part of this statement discusses this selection process further and highlights the importance of systematically collecting valid and robust animal-based measures. It is suggested that not only would this help a wide range of stakeholders benchmark results, but the availability of such data would help EFSA move towards quantitative animal welfare risk assessment.

The overall aim of the work outlined in this statement is to move towards establishing a common framework for future scientific opinions dealing with the use of animal-based measures to assess welfare. This statement is mainly intended therefore to support the work of EFSA, and a list of considerations for experts working on these future scientific opinions is given at the end of this statement. However, it is also hoped that this statement will be of interest to others working in the area of animal welfare risk assessment or in animal welfare assessment in practice.

1.1. EFSA scientific opinions on animal welfare and research on animal welfare assessment

Since 2004, EFSA has been asked to provide scientific opinions on the welfare of different farmed species, such as, laying hens, calves, pigs, dairy cows, broilers etc. The resulting scientific opinions (e.g. EFSA, 2004, 2005a, b, 2006, 2007a, b, c, 2009a, b, c, d, e, 2010a, b etc.) consist of a scientific report, including all data available about the welfare of the considered species, a formal risk assessment whenever possible, conclusions and recommendations from the available data and, in some cases, the outcomes of the risk assessment. In the risk assessments, factors that are risks for poor welfare (hazards) were identified and risk factors were assessed separately depending on the animal categories and housing systems. Factors which have beneficial effects for the animals were also considered and, in the future, formal benefit assessments may be carried out. A common recommendation from these opinions was that the body of research on animal welfare should be incorporated into codes of practice and monitoring protocols that address factors having potential positive and negative effects and to incorporate animal-based measures of welfare outcomes.

The EFSA opinions were based on a multidimensional concept of welfare that includes both the physical and mental state of the animal. This reflects currently accepted definitions and statements related to animal welfare. For example, Broom (1986) defines the welfare of an individual as "its state as regards its attempts to cope with its environment". The Farm Animal Welfare Council of the UK

proposed that an animal's welfare should be considered in terms of "five freedoms" that define ideal states rather than standards for acceptable welfare. These are (1) freedom from hunger and thirst, (2) freedom from discomfort, (3) freedom from pain, injury or disease, (4) freedom from fear and distress and (5) freedom to express normal behaviour (FAWC, 2009). The World Organisation of Animal Health (OIE) describes animal welfare as how well an animal is coping with the conditions in which it lives and considers the welfare of an animal to be good if, as indicated by scientific evidence, it is (i) healthy, comfortable, well nourished, safe and able to express key aspects of behaviour and (ii) it is not suffering from unpleasant states such as pain, fear and distress. In addition, the OIE considers that good animal welfare require disease prevention and veterinary treatment for illness and injuries, appropriate shelter, management, nutrition, humane handling and humane slaughter/killing (OIE, 2011). Previous EFSA opinions also recognise that, although the term "animal welfare" refers to the state of an individual animal, in practice, measurements of welfare in individual animals are often used to assess welfare at the group level (i.e. flock, herd, etc.).

There is a considerable body of research related to the use of animal-based measures to assess the welfare of animals (see Smulders et Algers, 2009, and Appleby et al., 2011, for recent overviews). The largest initiative so far is the Welfare Quality® project, the overall aims of which were to use practical strategies to improve animal welfare by developing a standardised methodology to assess animal welfare and translate the assessments into easily understandable labelling information (Blokhuis et al., 2003). This project differed from the EFSA opinions in that it did not aim to identify factors resulting in good or poor welfare. The Welfare Quality® project focused primarily on animal-based measures that can be monitored and used during a single short farm visit by an independent inspector to assess current levels of welfare. These are therefore measures of a welfare outcome at a particular point in time.

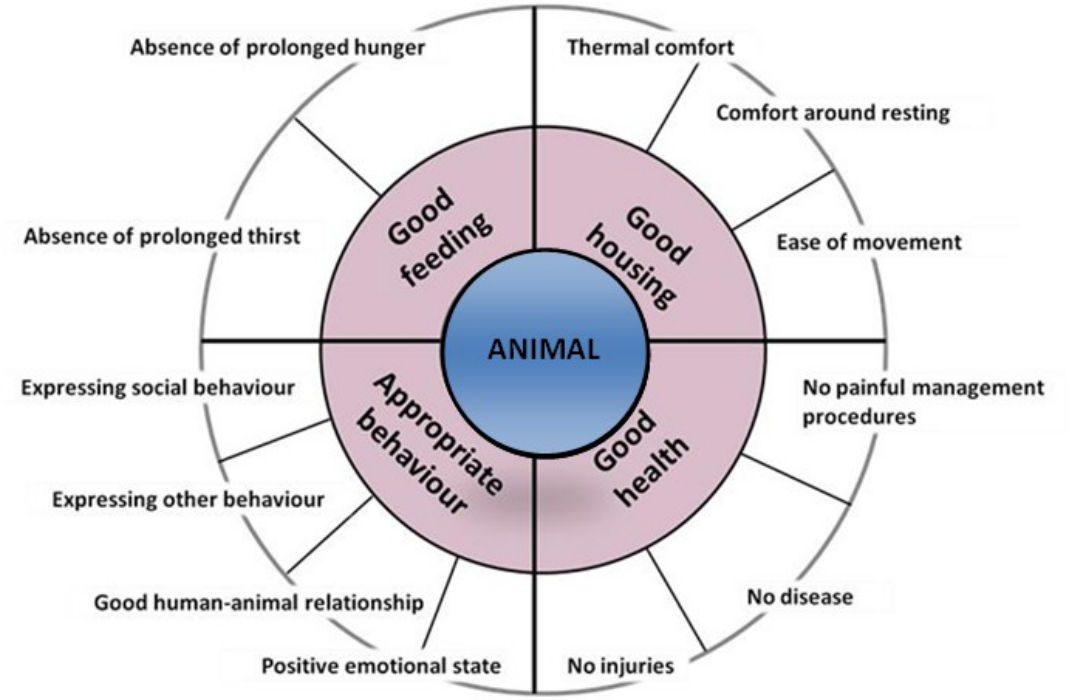


Figure 1: The four principles and 12 animal-based criteria used as guidelines for good welfare according to the Welfare Quality® project

Welfare Quality® proposes four welfare principles, linked to 12 criteria (Figure 1) that result in good welfare (Blokhuis et al., 2010) and can be considered a useful guideline for achieving it (Rushen et al., 2011). Again the multidimensional aspect of welfare is emphasised. The welfare criteria are in turn linked, in the detailed Welfare Quality® documents, to a series of measures, such as body condition, presence of injuries, etc., for cattle, pigs and poultry (Forkman and Keeling, 2009a, b, c; Welfare Quality®, 2009a, b, c). The principles are relevant to housed animals kept on land, and principally those on a farm, but were not designed to be comprehensive for welfare assessment in other species or circumstances, e.g. aquatic animals or wild animals, even though the same principles may apply.

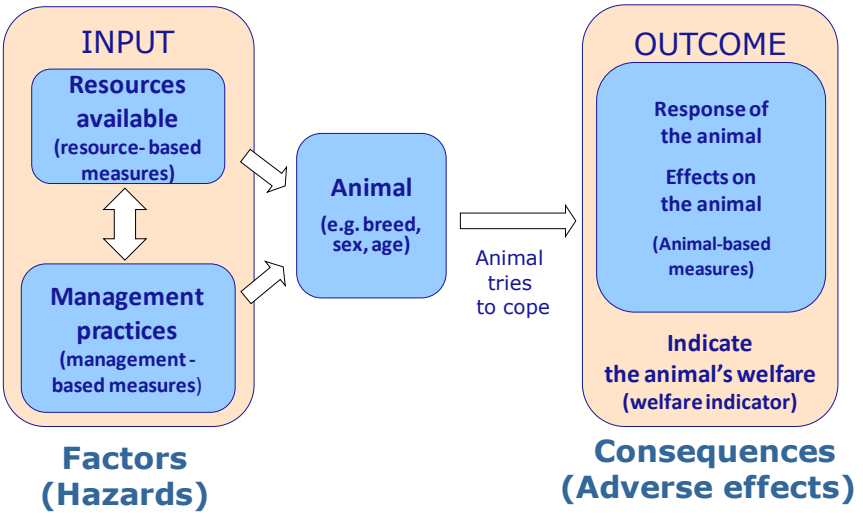
Blokhuis et al. (2010) proposed that, as animal welfare is a multidimensional concept, all criteria are important, so, for example, the ability to exhibit appropriate behaviour does not compensate for poor health, and good health does not compensate for behavioural problems. Botreau et al. (2009) proposed a formal model in which the above-mentioned measures presenting Welfare Quality® information could be transformed into value scores that express compliance with the 12 criteria and then with the four principles.

In general, the concepts of animal welfare used in the Welfare Quality® project overlap considerably with the EFSA scientific opinions referring to housed animals. The areas of welfare concern addressed in the EFSA Opinion on dairy cattle (EFSA, 2009a, b, c, d, e) were metabolic and reproductive disorders, udder disorders, leg and locomotion problems and behavioural disorders, fear and pain; in the EFSA and other EU scientific opinions on pigs (EFSA, 2004, 2005b, 2007a, b, c) they were castration, tail biting, different space allowances and other aspects of housing and husbandry systems for the various categories of pigs; and in the EFSA opinions on poultry (EFSA, 2010 a, b) they were the consequence of selection for fast growth in broilers, all of which are also addressed in the Welfare Quality® project. The main exception to this similarity is that Welfare Quality® included more signs of good welfare, i.e. positive emotional states, than the EFSA opinions, which were mandated to be focused on negative states in their risk assessments. In addition, the Welfare Quality® project specifically addressed the relationship between measures within and between the different welfare criteria, and in the EFSA scientific opinions the relationships between risk factors are discussed. Nevertheless, neither Welfare Quality® nor the EFSA opinions before 2011 systematically linked risk factors and their welfare consequences (see EFSA, 2012c). By linking these risk assessment and welfare assessment approaches and considering benefit assessment, this statement contributes to establishing a common framework for future EFSA scientific opinions dealing with the use of animal-based measures in the assessment of welfare. It links to the on-going work within EFSA on risk assessment in animal welfare, which has previously involved the development of guidelines (EFSA, 2012c), but now also includes work on the types of data needed for EFSA risk assessments and is aiming towards a greater harmonisation on modelling and terminology. The statement can also help to identify issues in the communication between decision makers and assessors of risks and benefits that need clarification from one or both sides and highlight how animal-based measures could be used to monitor animal welfare in Member States. The recently adopted EU Strategy for the Protection and Welfare of Animals 2012–2015 highlights that the possibility of using scientifically validated outcome-based indicators complementing prescriptive requirements in EU legislation will be considered when necessary. In this context, the issues analysed in this statement may help to further strengthen the dialogue between assessors of risks and benefits and decision makers and support the latter in achieving the set objectives.

1.2. Terminology and integration of concepts

The factors that affect an animal's welfare include the physical environment and resources available to the animal (**resource-based measures**), such as space allocation, housing facilities, bedding material, etc., and the management practices of the farm (**management-based measures**), such as how often cows are milked, whether or not anaesthetics and analgesics are used in mutilations, breeding strategies, etc. Of course, factors also interact with each other, thus influencing the way they act on the animal. In risk assessment, it is the negative consequences of the factors that are considered, and so the term "hazard" is used to mean a factor with the potential to cause poor welfare. Depending on its

characteristics (breed, sex, age, etc.) the animal will respond to these inputs and the animal's responses are assessed using **animal-based measures**. In risk assessment terminology these responses are the consequences of the factors acting upon the animal. If they are negative they are referred to as adverse effects and if positive they are referred to as benefits. Both factors and consequences can be characterised using appropriate resource, management and animal-based measures.



"Hazards" and "adverse effects" are used in earlier EFSA opinions

Figure 2: An overview of concepts and terminology

In Figure 2 the word "**measure**" is used to mean a form of evaluation rather than an intervention intended to deal with a problem. A **measurement** is the result of this evaluation, e.g. *x* centimetres of feed trough, *y* cm^2 as the area of a lesion on the body or *z* times per week as the frequency of a cleaning operation. Animal-based measures are evaluative, obtained in a precise way and usually quantitative. They give an indication of an animal's welfare, but a set of measures is normally needed to provide a good assessment of welfare. It is also important to make a distinction between the "measure" and the "method" used to measure it. In some cases, the measure is limited by its method.

Animal-based measures can be collected on-farm either by observation or inspection of the animal or by assessing the effects of a response on the environment, e.g. loose faeces on the floor is evidence of diarrhoea in the group, although further investigation may be necessary to identify the affected individual. Data can also be collected at a slaughterhouse during meat inspection, by the use of disease reporting systems (surveillance), or by consulting production records. For this reason a distinction is sometimes made between direct animal-based measures, taken from the animal, and indirect animal-based measures, e.g. taken from records or by remote monitoring of behaviour. Although animal-based measures can be collected at any time in the animal's life, they are obviously easier to collect at some times than others, e.g. there are discussions of how best to collect animal-based measures from animals kept extensively on ranches.

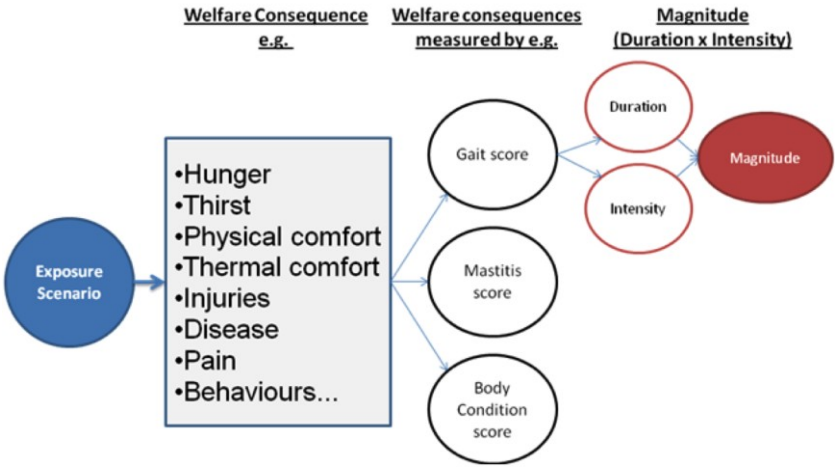
The terms described in this section are summarised below.

- Animal-based measures:
  - *Observations and measures* from the animals made during the welfare assessment on farm, ante or post mortem, e.g. behaviour, body condition (direct indicators), some of which involve veterinary procedures that can be carried out only by a veterinarian or other authorised individual (e.g. blood sampling).

- *Records* of animal breeding, growth, health, culling rate, abattoir condemnations, etc. These may include records of animal-based measures obtained using automated methods (indirect indicators, e.g. records of water consumption).

- Non- animal-based measures (resource- and management-based measures):
  - *Observations and measures* of housing provided or of management used (e.g. floor type, feeding space, weaning age or the use of legally permitted mutilations such as castration, tail docking).
  - *Documentation* (e.g. food provision strategies, staff training records, use of antibiotics).

In most cases the responses of an animal are adaptive, with little impact on its welfare, which indicates that an animal can cope rather easily with the factors to which it is exposed. But sometimes a response is of such a magnitude that it indicates the animal has had difficulty in coping or was not able to cope with these factors and, consequently, that the animal's welfare is impaired. Each animal-based measure will pertain to a single individual and to a particular time range. Sometimes a major response or effect can be the **cumulative outcome** of many days, weeks or months of more minor responses or effects, such as those which might follow abrasion against equipment, chronic stress or prolonged lack of appropriate nutrient levels in the feed. According to risk assessment terminology, the magnitude of a consequence is obtained by multiplying the duration of the consequence by its intensity. This was illustrated in the guidelines for risk assessment of animal welfare (EFSA, 2012c) (see Figure 3).



The red circle "duration" refers to the length of time for which an animal has had a particular gait score. The red circle "intensity" refers to the extent of the walking disorder, i.e. the actual score. The red circle "magnitude" refers to the overall consequence, lameness, and would need to be integrated with the magnitude of any other welfare consequences for an overall welfare assessment.

Figure 3: Consequence characterisation flow chart from the guidelines for risk assessment of animal welfare (EFSA, 2012c)

Expressing magnitude as a simple multiplicative function of maximum **intensity** and **duration** is probably too simplistic. In the laboratory animal science literature (e.g. LASA, 2008) and in the more recent EFSA animal welfare risk assessments (e.g. EFSA, 2012d) intensity is classified as mild, moderate or substantial and duration as short, medium or long. But even this does not take into consideration changes over time. For example, pain may be of a high intensity directly after an injury, but then gradually decrease as the wound heals, whereas pain attributable to an infection is likely to increase gradually and subside gradually over time. Broom (2001) proposed to represent this by a graph in which the intensity of the consequence (y-axis) is plotted against duration on the x-axis and where magnitude is then the area under the curve. This gives greater flexibility to follow the biological course of a welfare outcome over time and is therefore more likely to represent the true situation for the animal. The shape of the curve could be determined by repeated systematic collection of the animal-based measure used to quantify that welfare consequence. This approach becomes complex

when the magnitude of different welfare consequences, represented by different animal-based measures on different axes, are combined in multidimensional space.

Furthermore, although welfare is a characteristic of the individual animal, often a **group-level animal-based measure** is used (see Glossary). A simple example of this is the occurrence of lameness in a herd or flock. Gait scoring is a relatively well-studied animal-based measure of the pain felt by an individual animal when walking. A group-level measure that is derived from this gait scoring is the percentage of animals in a group that are currently above a certain level of lameness (Main et al., 2010). Sometimes new measures are generated by combining different animal-based measures from the same individual. An example of such a **composite animal-based measure** is the production record (milk yield, milk composition and calving interval) for an individual cow over three lactations, which can be used to assess the extent to which her initial productivity may have been compromised by the welfare consequences of factors such as nutrition, housing and management. Feed conversion ratio is another composite animal-based measure derived by combining feed consumed and growth rate. These factors, too, may be expressed as an average for the group.

Although many animal-based measures are simple and easy to use even under commercial conditions, in some cases the measure may require further analysis in a laboratory, e.g. metabolic profiling, or may be time-consuming to collect, e.g. changes in diurnal rhythm. However, with continued technical developments, especially those associated with automatic recording and precision livestock farming techniques, it is likely that several currently impracticable animal-based measures will become inexpensive and feasible in the future (Wathes et al., 2008).

Animal-based measures have usually been used to identify animals whose welfare is poor, but it is also desirable to identify as early as possible those animals whose welfare is deteriorating, so that changes can be made before the individual is further adversely affected. It may also be useful to be able to identify improvements in welfare as soon as possible in order to maximise benefits. Such measures could help to predict those animals at risk of poor welfare if no change or intervention is made and to promote good welfare where possible. An example of such an animal-based **predictive welfare indicator** is milk somatic cell count to indicate subclinical or clinical mastitis. For example, a high somatic cell count may not be a substantial welfare problem at the time of detection, but if steps are not taken it may become so, for example the animal may develop clinical mastitis. Similarly, changes in water consumption may predict later development of illness (Hegelund et al., 2003). Thus, in monitoring and surveillance systems, some animal-based measures may be identified as useful not because they indicate a current welfare problem, but because they are an indicator of the start of a cascade of potential negative welfare outcomes that are to be avoided. This cascade of decreasing welfare may be reflected in a gradual worsening of a single animal-based measure, i.e. an increase in intensity, to the point that it is an indicator of very poor welfare (e.g. body temperature), or it may be that the animal-based measure is a reliable predictor for the occurrence of another animal-based measure indicating a different serious response of the animal and hence of poor welfare (e.g. increased time lying may predict clinical lameness (Ito et al., 2009). In practice, clinical signs may be used as predictive indicators. But these, often slight, changes from normal are usually not quantified and there is a clear need for work in this area. In addition, the initial responses of the animal may feed back and modify the original factor or other factors in the environment, e.g. diarrhoea, may, of necessity, affect cleaning routines and hygiene (a management-based measure).

Input factors of the environment (non-animal-based measures) are often considered easier to record than the responses of the animal and they are more likely to be consistent over time. In some cases their effect on animal welfare is so well documented that they are a reliable proxy for animal-based measures. For example, the provision of water (a resource-based measure) is well correlated with the level of thirst. But the relationship between such environmental factors and animal-based responses is usually far from simple as different factors may interact to produce different responses in the animal and the response itself may vary according to characteristics of the animal, such as its age, stage of lactation, and so on.

The EFSA scientific opinions on the use of animal-based measures to assess the welfare of dairy cattle and pigs contain several examples of animal-based measures and also, where appropriate, non-animal-based measures (EFSA, 2012a, b). They were identified from the scientific literature as measures that could be used to determine the extent to which the recommendations in earlier EFSA opinions are fulfilled. However animal-based measures are used in research, as part of farm animal management, etc., and not only in the evaluation of welfare outcomes during an inspection visit. The following section gives a brief overview of these many potential uses of animal-based measures.

2. OVERVIEW OF THE USE OF ANIMAL-BASED MEASURES FOR WELFARE ASSESSMENT

The complete list of animal-based measures of welfare is long and it would be quite unrealistic and unnecessary to recruit all measures on every occasion that the welfare of an animal is to be assessed. They should be considered a comprehensive toolbox from which to select the range of measures necessary to address the specific objectives of a specific assessment for that particular species and category of animal at that time. That is to say the measures chosen should be "fit for purpose". For example, investigation of issues relating to overall welfare of animals on a farm (e.g. those that form part of an on-going health and welfare plan) requires that observations of animals be supported by records of performance, fertility and health. These are necessary because it is not possible to obtain a sufficient indication of welfare and the quality of husbandry on a farm from observations made during a short visit. Ability to prioritise welfare consequences for intervention purposes is critical but conditional on the development of an adequate database. In contrast, assessments that are very specific, for example on the impact of nutrition and feeding practices on the welfare of dairy cows, the impact of floor and space allowances on pig welfare, or the impact of genetic selection on broiler welfare, would require the selection from the toolbox of a narrower, but probably more focused, set of measures. An assessment made to check compliance with species-specific legislation would also focus only on those aspects covered by the legislation. A follow-up visit to an official inspection at which welfare problems had been identified, or a welfare assessment to support a ban on a particular management system, would need to provide legally reliable evidence, including that from veterinary procedures, which may not normally be obtained from a routine inspection.

Benchmarking is increasingly being used to track changes within the same farm over time as a part of good farm management. Comparison of the same measure between farms with similar housing systems or management practices facilitates the identification of those farms that are outside the normal range of variation in a specific animal-based measure. Such benchmarking of a harmonised set of standardised animal-based measures is especially useful to confirm improvements in welfare following a change. Benchmarking could also be used by breeding companies and legislators to track changes over time as a consequence of changes in selection criteria or following particular interventions or initiatives.

Below is a list of some of the potential users of animal-based measures.

- farmers, to support their management decisions or to track changes in welfare as a result of changes in management or environment
- consultants or advisers to farmers
- veterinary practitioners and veterinarians involved in herd health management
- breeding companies as part of their selection procedures
- auditing or accreditation organisations, to check that a farm satisfies the necessary criteria to be a part of a quality assurance or labelling scheme
- competent/responsible authorities, to check that a farm satisfies animal welfare requirements according to legislation and to evaluate effects in practice of changes in animal welfare legislation, or as part of pre-testing the welfare consequences of any future housing or technical development before it goes on the market
- scientists carrying out an experiment, so that their results can be compared with the results collected by other scientists.

2.1. Essential attributes of animal-based measures

In many respects the issues relevant to the animal-based measures used to assess animal welfare are similar to those used for any diagnostic test. Diagnostic tests are linked to disease-related outcomes, whereas animal-based measures are linked to welfare-related outcomes. We use these terms in specific contexts, but it is suggested that the criteria applied to diagnostic tests could also be applied to animal-based measures. The terms used in diagnostic testing as well as the terms commonly referred to in the animal welfare scientific literature are summarised below and subsequently used in the context of this statement.

To be appropriate and effective when used to assess welfare at farm level, animal-based measures need to be valid and robust. Validity might be situation dependent: sometimes measures that are valid during one stage of the production cycle are not applicable at other times. Specific requirements therefore need to be developed for each animal-based measure to ensure validity, such as how often and how long, at what time of the day, at what stage of lactation or pregnancy, etc., they should be applied. These are the same issues as those presented in the Welfare Quality® protocols and other animal welfare scientific literature, in which the more commonly used phrases are that a measure should be valid, reliable and feasible.

In the OIE diagnostic test validation guidelines (OIE, 2009), the fitness of a test encompasses being both fit for the *intended purpose* and fit for the *intended use*. **Fit for purpose** means that the test methods and related procedures must be appropriate (properly validated) for the specific purpose (such as screening blood samples in the context of herd freedom from disease certification). **Fit for use** takes into account practical aspects that might impact the relevance of the test or assay with respect to its intended application, e.g. its acceptability by scientific and regulatory communities, feasibility, and so on.

In the context of animal welfare assessment, the **appropriateness** (fitness for purpose) of an animal-based measure would relate to (a) how well it correctly assesses a specific welfare outcome (consequence) and (b) how practical it is. The appropriateness of a whole welfare assessment protocol would relate to how well the combination of measures reflects the overall welfare of the animal and how practical the protocol is when done in the field. Assuming the welfare assessment protocol consists only of animal-based measures that are themselves fit for the purpose of measuring a specific outcome (i.e. valid), the key issue for an assessment is which measures are to be combined. Even if the aim is not an overall assessment, but much more specific, the appropriateness of the welfare assessment still depends on which measures are taken.

Validation, in diagnostic tests, includes estimates of the analytical (biochemical) and diagnostic (population-based) performance characteristics of a test. In the context of this statement the focus is on the diagnostic properties, i.e. diagnostic **sensitivity** and **specificity**. In combination these are sometimes referred to as **accuracy**, which in a welfare context would be similar to the overall correctness of a measure to identify a specific welfare outcome. As with diagnostic tests, insensitive or non-specific indicators can potentially be useful for identification of welfare problems, especially when collected repeatedly or on a larger number of individuals. The use of screening tests is well developed, and some animal industries use their production databases to screen for farms with potential welfare problems. Non-specific indicators may satisfy some of the criteria for so-called "iceberg indicators". An iceberg indicator provides an overall assessment of welfare, just as the protruding tip of an iceberg signals its submerged bulk beneath the water's surface (FAWC, 2009). The example given in the FAWC report is the presence of an intact uninjured tail on a growing pig at slaughter. This would indicate that the tail has been neither docked nor bitten, with the implication that the animal's husbandry and management were of high quality. Other measures are sufficiently sensitive to detect changes that are within the normal biological range of an animal's responses, such as minor changes in number of heart beats per minute, which may or may not be relevant as welfare outcome indicators. Many things will determine the appropriate level of sensitivity required in a particular welfare assessment. Specificity relates to whether the animal-based measure is related to a

single welfare consequence or whether it relates or responds to several different consequences. This point is discussed in more detail in section 3 of this statement.

Robustness is another essential attribute of an appropriate animal-based measure for the assessment of animal welfare. It influences how the measure is affected by changes in the environment, who is taking the measure and when it is taken. It encompasses concepts such as **repeatability** and **reliability**, which are the agreement between repeated measurements of the welfare consequence on the same sample by the same assessor (**intra-observer)** or a different assessor (**inter-observer)** respectively. Maintaining repeatability and reliability over time requires training at regular intervals so that observers are "recalibrated" to a reference standard for the measure. This is very important to promote harmonisation of recording to ensure consistency and accuracy of measurements. A word very similar in meaning, sometimes used in the diagnostic literature, is **reproducibility**, which is the ability of the test to provide consistent results. In a welfare context this would refer to the ability of the animal-based measure to be applied under various conditions, by different personnel, while still providing similar (correct) results.

Whenever welfare outcomes vary over time, for example if they vary according to time of day, or time interval since last feeding, then the **measures should be based on a representative time sample**. This is mostly true of behavioural measures; for example, measures of how much time cows spend lying down should be based on observations made throughout the day rather than occasional observations. Similarly, many physiological measures show diurnal variation and so need to be based on an adequate number of blood or tissue samples. Furthermore, measures which are valid at one part of the production cycle may not be applicable in other phases. For these reasons it is important to specify when the measure is to be taken. If the measure is affected by time, this should be taken into consideration in the interpretation. They should not be affected by external factors that are not related to the welfare of the animals. For example, if the welfare of the animals does not change with weather or time of year, then the measures should also not be affected by weather or time of the year. If they are thus affected, appropriate allowances should be made, or efforts made to ensure consistency

Whenever measures are taken from only a sample of all animals in the unit, **it is essential that the sample be unbiased and representative** (e.g. in terms of sex, age, body size, etc.). It is important to specify how this sample of animals is to be chosen and the number of animals in the sample. The use of good operating procedures and reporting standards developed in health research should be applied to *all* animal-based measures.

Table 1: Summary of criteria for valid and robust animal-based measures

Criteria	Explanations and examples
They should accurately measure and indicate the welfare consequence for an animal	There are several ways of assessing validity, such as expert opinion or (preferably) by deriving a study-based diagnostic validity related to the relationship between a specific welfare outcome indicator and an independently performed assessment of the welfare outcome
Where the measures vary over time, e.g. time of day or interval since a particular event, then the measures should be based on a representative time sample.	This is particularly true for behavioural measures, e.g. how much time animals spend lying down. Furthermore, indicators which are valid at one part of the production cycle may not be applicable in other phases
They should have low variability when repeatedly measured by the same observer	This means low intra-observer variability and resulting high repeatability
They should be consistent when measured by different observers on the same animal	This means low inter-observer variability and high reproducibility. People should be trained to the "gold standard" for the measure and this training should be repeated at regular intervals so that observers are "recalibrated"
They should not be affected by external factors that are not related to the welfare of the animals	If the welfare of the animals does not change with weather or time of year, then the measures should also not be affected by weather or time of the year. This indicates a high robustness
Taking the measures should be feasible for the purpose of the data collection	They should not be costly to make and should not involve much observer/farmer time, making them practical and feasible

Animals may be inspected either at the farm or during ante-mortem or post-mortem inspection in the slaughterhouse. Animal-based measures taken during ante-mortem inspection that provide information about welfare on-farm include identifying severe lameness, injuries, clinical disease or starvation, as indicated by body condition. There are other ante-mortem and post-mortem slaughterhouse indicators that give information about welfare during transport, lairage and pre-slaughter handling (e.g. injuries, fear reactions). Such measures are more commonly used in animals kept for meat, e.g. fattening pigs, beef cattle and broiler chickens. Information relevant to the assessment of welfare during meat inspection is available in the recent EFSA opinions on the topic (EFSA, 2011).

From the previous sections, it is clear that there are potentially many different animal-based measures that can be used to assess animal welfare. Which measure is the most appropriate for a particular situation will therefore depend on a number of different issues, e.g. the purpose of the assessment, the skills of the person collecting the measure, the conditions under which it is to be gathered and the time available to collect it, as well as financial constraints. Knowing how different animal-based measures vary on these different attributes as well as what methods can be used to collect them is important as these factors influence whether a particular measure is appropriate (fit) for a specific purpose. At present, the potential and limitations are known for some animal-based measures, as are some of the important considerations to be taken in consideration when selecting the methods for recording such measurements, but this is not the case for all measures.

3. DEVELOPMENT OF TOOLS TO MONITOR ANIMAL WELFARE

A proposal in this statement is that the essential characteristics of the measures should be described, as well as their potential uses and limitations, so making it easier for a wider range of end users to select the most appropriate measures for different purposes. There are certain basic similarities in how this system would work and health monitoring systems (Salman, 2003). A specific example related to lameness is given below (Figure 4), but in general the system is made up of the following steps. The first step is identification of the goal. The second step is the identification of the population concerned and the definition and selection of the survey population. The third step is the selection of the animal-based measure, or combinations of measures, from the "toolbox" and the systematic collection of data. Following the analyses of the data, the results are interpreted. This analysis and interpretation may in practice be done automatically as part of management software or on the website of the appropriate animal industry. In some cases a recommendation for action is developed and implemented. The goal and the survey population are reappraised and when necessary adapted and then more data collected on the same measure(s) to verify whether the action has resulted in the intended effect.

In a simple case, a farmer or advisor whose goal is to reduce lameness in dairy cows on a farm would select from the toolbox the most appropriate animal-based measures of lameness, perhaps a gait-scoring system, and start to collect data on the gait score of the dairy cows on the farm. This would provide reliable information about the prevalence and magnitude of the lameness problem on the farm, perhaps even allowing benchmarking against other farms. The farmer or advisor may also decide to keep records from the routine hoof-care treatments. If hazards were also identified on that farm and recommendations developed or followed to reduce those, e.g. change the floor or improve manure removal, then it would be useful after this action to collect data on gait scoring again and to study the hoof-care records over time to see if the welfare problem had been reduced, and so on, until the farmer/advisor was satisfied that the goal of reducing lameness on the farm had been achieved. At this point, either monitoring is stopped or the programme is redirected towards a monitoring of maintenance and demonstration of the good status achieved. This is visually represented in Figure 4.

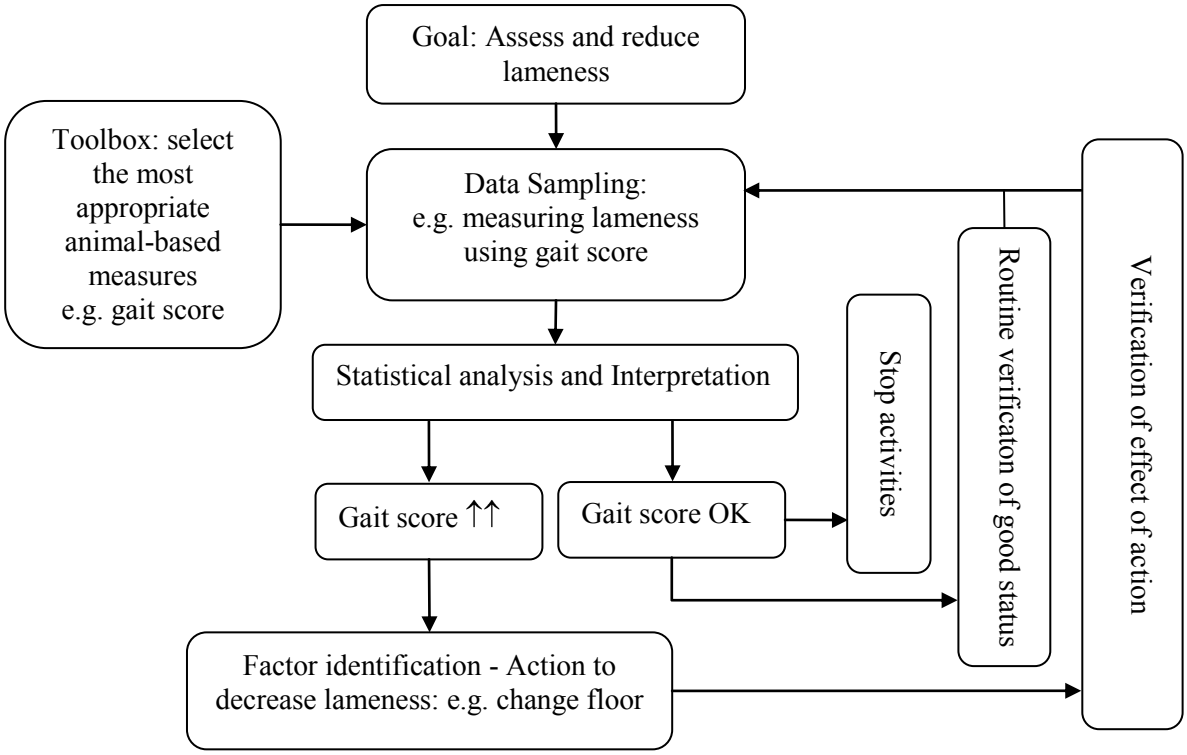


Figure 4: Schematic representation of steps involved in monitoring of animal welfare using lameness as an example

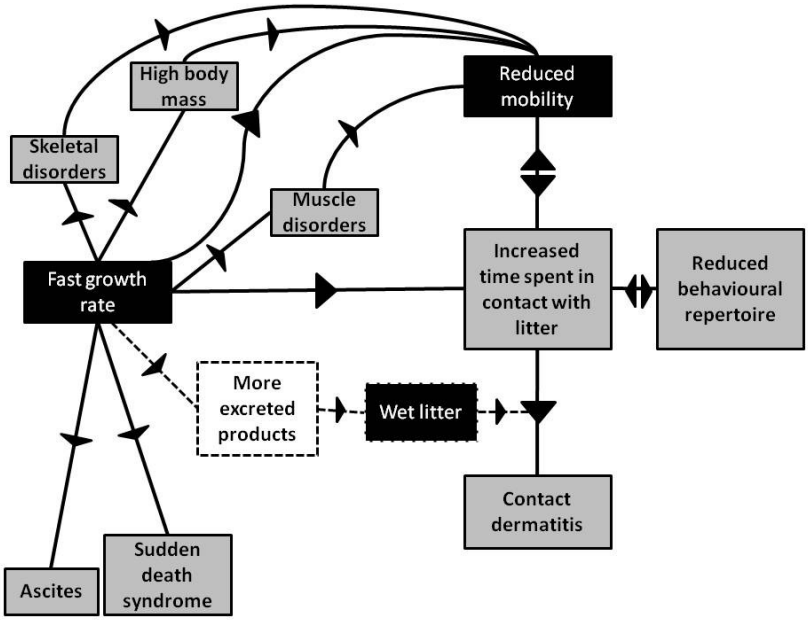
Another example may be that of a legislator with the overall goal of evaluating whether changes in the animal welfare legislation are improving welfare. In this case, the legislator could select a combination of welfare outcome indicators that cover all aspects of welfare and collect data on these. This would be the welfare equivalent of a general health monitoring programme. More common, however, in the area of animal health would be to focus on specific health issues or diseases, e.g. as in syndromic monitoring. Thus, it is also possible, and probably easier, for a legislator to choose to focus on a few main welfare issues as subgoals. For the purposes of this example we can consider the behavioural problem of tail biting (EFSA, 2007c). The next step for the legislator, in this example of tail biting, would be to define the target population (EFSA, 2012c) and to select from the toolbox the most appropriate animal-based measures for this consequence. In this example, the most appropriate animal-based measure is probably a tail damage score. Tail damage scoring may be carried out on farms, although, as it is most common in fattening pigs, it may be easier to standardise at slaughterhouses. When repeated on subsequent occasions, the legislator may be able to identify trends, i.e. a general increase or decrease in tail-biting behaviour over time, or compare different subset of data, e.g. from different housing systems or regions. Alternatively, the data may be used to set target levels for tail damage and realistic dates by which these targets can be achieved.

The process by which EFSA may select animal-based measures for use in future scientific opinions is discussed in section 3.4 of this statement.

3.1. Visualising links between factors and animal-based measures

When only a single welfare outcome (consequence), being the consequence of a single existing or acting input factor (hazard), is targeted, and this outcome can be measured (detected) by a single well-known animal-based measure (welfare indicator), then a few well-designed studies can establish the strengths of existing associations between the three layers, i.e. the measure (the indicator or test), the outcome (welfare consequence) and the factor (hazard). This information could then be retrieved from the scientific literature (study publications). However, the system becomes very complex when envisaging a situation with (a) several non-specific animal-based measures that are linked to a given welfare outcome, (b) several factors that affect the same welfare outcome and (c) associations at the level of (i.e. between) measures, welfare outcomes and factors. Quantification of the large number of within- and between-layer associations in such a multilayer web requires the systematic collection and analysis of a broad range of data from various sources. This quantification, however, is necessary in order to select the optimal animal-based measure (or a combination of several measures) from the toolbox for a given question. This became very clear when developing the scientific opinions on animal-based measures trying to address how animal-based measures could be used to ensure the fulfilment of the recommendations of the EFSA scientific opinions and how the assessment protocols of the Welfare Quality® project cover the main hazards identified in EFSA scientific opinions and vice versa.

The process showed that there are multiple interactions between animal-based measures. For example, a lame cow may be less competitive at the food trough, and so not have the most appropriate diet, therefore increasing its risk of metabolic disease. At the same time, it may lie for longer periods of time, so increasing its risk of mastitis if hygiene in the stall is not optimal on that particular farm. However, assuming that following a consulting process a combination of measures was finally chosen, information gathered in a well-structured approach could contribute significantly to visualising these links, thereby contributing knowledge when future decisions about optimal combinations of animal-based measures are to be made. For example, it may be unnecessary in the future to record all the selected measures if there is some overlap in the information they contain. In addition, if information on some aspects of the farms being monitored is available, e.g. the housing system or even more detailed information, it becomes feasible to visualise links between such factors and animal-based measures. This complexity was illustrated in the EFSA scientific opinion on the influence of genetic parameters on the welfare and the resistance to stress of commercial broilers (EFSA, 2010a), in which it was shown that fast growth is linked both directly and indirectly to other hazards in the risk assessment, which are considered as independent factors (see Figure 5).



Fast growth rate is shown to be linked both directly and indirectly to other hazards characterised in the risk assessment, which are considered as independent factors. Black boxes are hazards characterised in the risk assessment. Grey boxes are hazard consequences. White boxes explain the relationship between hazards where necessary. Arrows show the direction of causality. Note that this image does not contain all the possible consequences of all the hazards shown.

Figure 5: The inter-relationships between hazards and consequences, using fast growth rate as an example starting hazard

These examples show that links between hazards, links between consequences and links between animal-based measures are not singular. Nor do they need to be of the same strength, i.e. of similar specificity. In several cases, different hazards might lead to the same welfare outcome, i.e. to the same consequence. That the situation is complex is already well known from risk assessment studies for infectious diseases. However, when considering consequences relevant to animal welfare we need to add to infectious diseases not only physical injuries to animals, but also behaviour disturbances, i.e. the whole spectrum of physical and mental health consequences for the animal. We should not restrict ourselves to infectious health consequences, as has been the tendency in the past.

3.2. Quantification of the strengths of links, of predictive and classification capacity

Welfare Quality® protocols and the EFSA scientific opinions on animal welfare can help when selecting appropriate animal-based measures to detect the presence of welfare hazards and to generally monitor animal welfare. However, description, analysis and visualisation of the full system will aid in the selecting the most appropriate tools for monitoring animal welfare. The work done on the scientific opinions on animal-based measures has shown that, owing to the complexity of the system and heterogeneity of unlinked data sources, it is currently difficult to combine information originally collected for different aims into a complete picture.

Furthermore, establishing only the links, but not the predictive capacities, of the animal-based measures or their combinations still does not allow selection of the most effective combination of measures for a specific goal. For example, it would be very helpful to optimise the toolbox so that we would know which combination of measures is best suited to and most efficient in measuring the presence of welfare outcomes and factors of interest. To fully exploit the potential of animal-based measures, combinations of measures need to be identified that are both sensitive and specific for each relevant welfare outcome and factor. Using the terminology developed in the Welfare Quality® project, this would be referred to as a valid measure since it really tells about some aspect of the actual welfare of the animal (Welfare Quality®, 2009a, b, c). To achieve this, one needs to identify and fully

explore the presence and the predictive capacity of the correlations or associations within the factor–outcome–measures network. There are three main approaches to achieve this: systematic reviews, expert elicitation and analysis of information extracted from databases.

Systematic reviews can be carried out only on very focused questions, and may include meta-analyses to quantify associations, which require that the information was collected through several well-designed studies and subsequently published. This is often very time-consuming. The **expert elicitation approach**, which is used for broader questions, is also limited by the time and resources available to "score" the potentially large number of paired links. It requires experts to extract the relevant information for each link from the existing (scientific) literature and to fill in the probably numerous data gaps with expert opinion. Issues related to (a) the selection of experts and (b) the lack of transparency in the final risk assessment are further disadvantages of the expert elicitation approach. The **database approach** is currently limited by the lack of systematically collected field data at the animal, group and farm level, captured in a centralised database, from which to explore interactions between factors, welfare outcomes and measures using specific statistical tools. An advantage of the database approach would be the improved transparency and consistency of results based on "objective data" and the possibility of moving towards quantitative risk assessment in animal welfare.

In order to further explore a possible route to proceed towards quantitative risk assessment of animal welfare, a report (Presi and Reist, 2011) was commissioned from a consulting company (Sanisys SA; www.sanisys.net). Sanisys was asked specifically to propose options for a suitable system to:

- systematically record (collect) relevant information in a user-friendly database, including both the existing data as summarised in the Welfare Quality® reports and several EFSA opinions as well as other published information, and information recorded from on-going studies or monitoring systems in the field;
- specifically analyse and visualise the links between and within the three levels (factors, outcomes, measures) with a single database system.

In the field of social sciences and network analysis, statistical methods have been developed and applied to identify and describe complex associations between elements in populations or networks. Increasingly such methods are also employed in animal science, for example to describe animal movements in populations, and thus to identify direct contact structures relevant in the context of welfare in general (Cañon Jones et al., 2011) or infectious disease outbreaks. In order to link several factors to a specific outcome, multivariable modelling approaches such as logistic regression analysis, log-linear models, principal component analysis, discriminant analysis (without or with variable selection) or classification (decision) trees can be used. Selecting the best approach will depend on the specific question to be addressed as well as the data format and structure. For rapid visualisation of the complex network structure, once the nodes and links have been identified, a range of commercial and open-source software packages (tools) are available.

Data compilation into one or more relational databases can be from one or several sources, e.g. on-going recordings (field records and monitoring), other databases, designated research projects, risk assessments and expert opinion. Automated data analysis routines can be used to facilitate communication between the different sources of information, to analyse the data and to extract appropriate information in the format of standardised reports. These reports can summarise the prevalence or incidence of factors and welfare outcome indicators and to benchmark the results. Data analysis can also contribute to identifying links and the strengths of the links between input factors and welfare consequences that are currently lacking because the complexity of the factor–outcome–measure network. When populated and used appropriately this database approach could assist in selecting the most effective animal-based measures from the toolbox, and would ultimately provide the type of information required for quantitative risk assessment of animal welfare.

3.3. Risk-based surveillance (monitoring plus corrective action/control)

Tools developed to monitor animal welfare need to deal with not only what is recorded and how it is analysed to generate new knowledge that can be used in risk assessment, but also the implications of the results gathered on-farm. Some aspects of this have already been mentioned already.

It may be that the aim is merely to monitor animal welfare, e.g. when deciding whether or not a product from a farm (meat or milk) can be included under a particular market label, and so help consumers choose animal welfare-friendly products. However, in many cases, the reason for monitoring implies that action will be taken as a result of the monitoring. If actions are foreseen when predefined threshold levels are exceeded, one typically talks about surveillance rather than monitoring. It may even be that the welfare status is monitored again, after the intervention, to determine whether or not there has been any improvement in welfare. Such a "before versus after" assessment would be necessary for economic analyses, for example when comparing the cost-effectiveness of different interventions with the gains in welfare post intervention. Such information is useful at the level of the farm, when making management decisions, but it would also be useful information at the country and EU level, when companies or governments are making policy decisions. The systematic recording of standardised animal-based measures from the validated measures in the "toolbox"' then becomes part of an animal welfare surveillance scheme.

Benchmarking of animal-based measures on a large scale might be particularly important for early detection of welfare changes that would not otherwise be detected, or would not have been detected until much later. This would allow the earlier detection of any potential problems leading to poor welfare as a result of trends in the sector, e.g. changes in breeding goals, changes in raw ingredients in feed, etc. On the positive side, benchmarking of important animal-based measures on a large scale would give quicker feedback to policy makers on the effectiveness of legislation or other initiatives to improve animal welfare.

Surveillance of outcome measures is already established in other areas and there are similarities between what is discussed here and diagnosis in animal health based on clinical signs and meat inspection.

3.4. Considerations for the future EFSA scientific opinions addressing the use of animal-based measures to assess the welfare of animals

As explained above, ideally there would be a database available for the species in question where resource-based, management-based and animal-based measures are all recorded. The process to develop the EFSA scientific opinion would then consist of identifying the most appropriate combination of animal-based measures to cover the factors known to influence welfare. In short, it would involve a quantitative risk assessment of the main factors (probably hazards) and the most efficient tools (probably animal-based measures) to monitor their (negative) consequences.

Unfortunately, as also stated previously, such databases do not yet exist. Thus, until such time as they do, the following steps are proposed.

1) Identification of the animal-based measures to be placed in the "toolbox" for this species.

    a. This could be done in many ways, but given that, for most species, there are already EFSA scientific opinions which have identified the *factors* known to affect welfare, it is proposed that the animal-based measures linked to these factors are identified from the scientific literature. As a confirmation that all factors are covered it is proposed to systematically address each *recommendation* from the previous EFSA scientific opinion and list which animal-based measure(s) could be used to determine whether or not that particular recommendation has been fulfilled. Since recommendations in EFSA scientific opinions are based on a factor (hazard) identification process, this should lead to the same list of animal-based measures.

b. At this stage any gaps, i.e. factors or recommendations for which there is no animal-based measure, will be identified.

2) Critical examination of the essential characteristics of the animal-based measures in the toolbox.

a. At this stage the validity and robustness of each animal-based measure is evaluated. Although no measure should be include in the toolbox that does *not* satisfy the criteria outlined previously in this statement, it is at this stage that the animal-based measures are assessed for their fitness for the purpose of the assessment. A valid and robust measure may be rejected at this stage because it is too time consuming or costly to collect for the intended purpose of the assessment.

3) Selection of a shortlist of animal-based measures for the purposes of the welfare assessment since it not possible and probably unnecessary to use all measures from the toolbox.

a. This could be achieved by selecting those animal-based measures that are considered most often to address a factor or recommendation. This would presumably lead to the shortest list, but these measures may not be specific to particular factors and so less useful if the purpose of the assessment is to recommend a change, by managing the causal factor, rather than merely monitoring the presence or otherwise of its consequences. Another weakness of this approach is that a high frequency of reference to a welfare problem does not necessarily indicate high impact on welfare.

b. Only scientific considerations should be used to determine which animal-based measures are placed in the "toolbox"'. But decisions on a shortlist of measures that are both comprehensive as well as acceptable to end users are probably best made in consultation with a diverse group of stakeholders.

c. At this stage consideration may be given as to whether a resource or management-based measure is sufficiently well correlated with an animal-based measure, but so much more feasible that it may be recorded instead of the actual animal-based measure.

d. Once the shortlist of animal-based measures (perhaps complemented by resource- or management-based measures) has been identified, a final check should be made to ensure that the measures do indeed cover the main factors and recommendations identified previously as important for the welfare of that particular species or category of animal.

4) Promote the establishment of a database for future quantitative risk assessment of animal welfare and identify a set of resource, management and animal-based measures that should be recorded for this purpose.

a. If, following the EFSA scientific opinion, the shortlist of animal-based measures is to be collected, then consideration should be given, even in the scientific opinion itself, of how risk managers might assemble the data so that they contribute in the best way possible to the establishment of the database proposed earlier. Although it is unlikely to be the comprehensive database necessary for a complete quantitative risk assessment, it will nevertheless provide valuable information for EFSA future work. Not least it will, even after a short period of time, allow an evaluation of the correlations and links between the currently collected animal-based measures and perhaps a refinement of the existing shortlist of measures.

b. In an ideal situation, stakeholders contributing to the establishment of this science-based information would compile the collected information in a standardised format, and the information would be entered into a central database (or be in some way electronically accessible for data retrieval through predefined interfaces). Over time, that compiled information would enable systematic analysis of the existing data, to quantify links between animal-based measures, welfare outcomes and factors, to provide visualisation tools for easier interpretation and selection of measures, and for gap analysis, i.e. to identify research needs and inform the type of data to be collected during research.

5) Emphasise the use of good operating procedures and reporting standards

a. A range of good operating procedures has been developed for the design, implementation and analysis of experimental and observational studies in health research. Given that there is no conceptual difference between studies designed to assess the association between a given (risk) factor and a disease-related outcome and studies looking at welfare-related outcomes, these good operating procedures are also of relevance in welfare research.

b. In addition to good operating procedures, reporting standards for clinical trials (CONSORT), observational epidemiological studies (STROBE) and diagnostic tests (STARD) and other health research reporting guidelines as summarised within the Equator network (http://www.equator-network.org/resource-centre/library-of-health-research-reporting/) should be used to present the results from welfare-related research.

c. Adherence to both good operating procedures and reporting guidelines substantially increases the potential of extracting, evaluating and utilising the relevant information from publications for use in risk assessments related to animal welfare.

CONCLUSIONS AND RECOMMENDATIONS

CONCLUSIONS

Animal-based measures are the most appropriate indicators of animal welfare and a carefully selected combination of animal-based measures can be used to assess the welfare of a target population in a valid and robust way.

The most appropriate combination of animal-based measures will depend on the purpose of the welfare assessment. In some situations, feasibility and economic aspects may be most important, whereas in others sensitivity and specificity may take priority.

To be able to select the most appropriate animal-based measures, more information is needed about the essential characteristics of the different measures (their accuracy, sensitivity, specificity, reliability, repeatability, etc.).

In order to ensure good welfare in animals, we need both to reduce the risk of poor welfare, by using resource-based and management-based measures, and to monitor actual welfare, by using animal-based measures.

More information is needed about the links between input factors and animal-based measures of welfare outcomes. In particular, the strength of the link and its predictive ability are important. This is necessary to be able to select the most appropriate combination of measures.

The systematic collection of standardised field data on animal-based measures and subsequent storage in well-defined databases can in the future assist in better assessing the validity and robustness of animal-based measures and thus the move towards quantitative risk assessment of animal welfare. Such data are also a potentially valuable tool to monitor animal welfare in Members States.

RECOMMENDATIONS

The steps outlined in this document should be followed in future EFSA scientific opinions on the use of animal-based measures in the assessment of animal welfare.

Those developing animal-based measures should use good operating procedures and reporting standards so that the data they gather can contribute to knowledge about the characteristics of the animal-based measure and the methods used to collect it.

Both animal-based measures and input factors (resource- and management-based measures) should be used in combination when monitoring animal welfare.

Although a harmonised method of recording a particular animal-based measure would be best, reaching agreement on this method may not always be possible. If more than one method of recording a particular animal-based measure is available, then there should at least be equivalence between the different methodologies.

How best to share and combine data when both input factors and animal-based measures have been recorded should be considered. More widely available data of this type would facilitate analyses to determine the strengths of the links between factors and consequences.

Once a long list of valid and robust animal-based measures has been identified in an EFSA scientific opinion on the use of animal-based measures to assess the welfare of a particular species, a shortlist should be defined in consultation with a diverse group of stakeholders.

There should be on-going evaluation of animal-based measures and how they are used to assess the welfare of animals as this is rapidly expanding area and new measures and new methods of recording established measures are being developed.

REFERENCES

Appleby MC, Mench JA, Olsson A and Hughes BO (Eds), 2011. Animal welfare, 2nd edn. CABI Publishing, Wallingford, 344 pp.

Blokhuis HJ, Jones RB, Geers R, Miele M and Veissier I, 2003. Measuring and monitoring animal welfare: transparency in the food product quality chain. Animal Welfare, 12, 445–455.

Blokhuis HJ, Veissier I, Miele M and Jones B, 2010. The Welfare Quality® project and beyond: safeguarding farm animal well-being. Acta Agriculturae Scandinavica Section A—Animal Science, 60, 129–140.

Botreau R, Veissier I and Perny P, 2009. Overall assessment of animal welfare: strategy adopted in Welfare Quality®. Animal Welfare, 18, 363-370.

Broom DM, 1986. Indicators of poor welfare. The British Veterinary Journal, 142, 524–526.

Broom DM, 2001. Coping, stress and welfare. In: Coping with challenge: welfare in animals including humans. Ed. Broom DM. Dahlem University Press, Berlin, 1–9.

Cañon Jones HA, Noble C, Damsgård B and Pearce GP, 2011. Social network analysis of the behavioural interactions that influence the development of fin damage in Atlantic salmon parr (Salmo salar) held at different stocking densities. Applied Animal Behaviour Science, 133, 117-126.

EFSA (European Food Safety Authority), 2004. Opinion of the Scientific Panel on Animal Health and Welfare on a request from the Commission related to welfare aspects of castration of piglets. The EFSA Journal, 91, 1–18.

EFSA (European Food Safety Authority), 2005a. Scientific Opinion of the Panel on Animal Health and welfare on a request from the Commission related to the welfare aspects of various systems of keeping laying hens. The EFSA Journal, 197, 1–23.

EFSA (European Food Safety Authority), 2005b. Opinion of the Scientific Panel on Animal Health and Welfare on a request from the Commission related to welfare of weaners and rearing pigs: effects of different space allowances and floor types. The EFSA Journal, 268, 1–19.

EFSA (European Food Safety Authority), 2006. Scientific Opinion of the Panel on Animal Health and Welfare on a request from the European Commission on the risks of poor welfare in intensive calf farming systems. An update of the Scientific Veterinary Committee Report on the Welfare of Calves. The EFSA Journal, 366, 1–36.

EFSA (European Food Safety Authority), 2007a. Scientific Opinion of the Panel on Animal Health and Welfare on a request from the Commission on Animal health and welfare aspects of different housing and husbandry systems for adult breeding boars, pregnant, farrowing sows and unweaned piglets. The EFSA Journal, 572, 1–13.

EFSA (European Food Safety Authority), 2007b. Scientific Opinion of the Panel on Animal Health and Welfare on a request from the Commission on Animal health and welfare in fattening pigs in relation to housing and husbandry. The EFSA Journal, 564, 1–14.

EFSA (European Food Safety Authority), 2007c. Scientific Opinion of the Panel on Animal Health and Welfare on a request from the Commission on the risks associated with tail biting in pigs and possible means to reduce the need for tail docking considering the different housing and husbandry systems. The EFSA Journal, 611, 1–13.

EFSA (European Food Safety Authority), 2009a. Scientific Opinion of the Panel on Animal Health and Welfare on a request from the Commission on the risk assessment of the impact of housing, nutrition and feeding, management and genetic selection on behaviour, fear and pain problems in dairy cows. The EFSA Journal, 1139, 1–68.

EFSA (European Food Safety Authority), 2009b. Scientific Opinion of the Panel on Animal Health and Welfare on a request from the Commission on the risk assessment of the impact of housing,

nutrition and feeding, management and genetic selection on metabolic and reproductive problems in dairy cows. The EFSA Journal, 1140, 1–75.

EFSA (European Food Safety Authority), 2009c. Scientific Opinion of the Panel on Animal Health and Welfare on a request from the Commission on the risk assessment of the impact of housing, nutrition and feeding, management and genetic selection on udder problems in dairy cows. The EFSA Journal, 1141, 1–60.

EFSA (European Food Safety Authority), 2009d. Scientific Opinion of the Panel on Animal Health and Welfare on a request from the Commission on the risk assessment of the impact of housing, nutrition and feeding, management and genetic selection on leg and locomotion problems in dairy cows. The EFSA Journal, 1142, 1–57.

EFSA (European Food Safety Authority), 2009e. Scientific Opinion of the Panel on Animal Health and Welfare on a request from European Commission on the overall effects of farming systems on dairy cow welfare and disease. The EFSA Journal, 1143, 1–38.

EFSA Panel on Animal Health and Welfare (AHAW), 2010a. Scientific Opinion on the influence of genetic parameters on the welfare and the resistance to stress of commercial broilers. EFSA Journal, 8(7):1666, 82 pp.

EFSA Panel on Animal Health and Welfare (AHAW), 2010b. Scientific Opinion on welfare aspects of the management and housing of the grand-parent and parent stocks raised and kept for breeding purposes. EFSA Journal, 8(7):1667, 81 pp.

EFSA Panels on Biological Hazards (BIOHAZ), on Contaminants in the Food Chain (CONTAM), and on Animal Health and Welfare (AHAW), 2011. Scientific Opinion on the public health hazards to be covered by inspection of meat (swine). EFSA Journal, 9(10):2351, 198 pp.

EFSA Panel on Animal Health and Welfare (AHAW), 2012a. Scientific Opinion on the use of animal-based measures to assess welfare of dairy cows. EFSA Journal, 10(1):2554, 81 pp.

EFSA Panel on Animal Health and Welfare (AHAW), 2012b. Scientific Opinion on the use of animal-based measures to assess welfare in pigs. EFSA Journal, 10(1):2512, 85 pp.

EFSA Panel on Animal Health and Welfare (AHAW), 2012c. Guidance on risk assessment for animal welfare. EFSA Journal, 10(1):2513, 30 pp.

EFSA Panel on Animal Health and Welfare (AHAW), 2012d. Scientific Opinion on the welfare of cattle kept for beef production and the welfare in intensive calf farming systems. EFSA Journal, 10(5):2669, 166 pp.

FAWC (Farm Animal Welfare Council), 2009. Farm Animal Welfare in Great Britain: Past, Present and Future. 57 pp. Available from HTTP://WWW.FAWC.ORG.UK/REPORTS.HTM

Forkman B and Keeling LJ (Eds), 2009a. Assessment of animal welfare measures for dairy cattle, beef bulls and veal calves. Welfare Quality® Reports No 11, 297 pp.

Forkman B and Keeling LJ (Eds), 2009b. Assessment of animal welfare measures for sows, piglets and fattening pigs. Welfare Quality® Reports No 10, 284 pp.

Forkman B and Keeling LJ (Eds), 2009c. Assessment of animal welfare measures for poultry. Welfare Quality® Reports No 9, 176 pp.

Hegelund L, Sorensen JT and Johansen N, 2003. Developing a welfare assessment system for use in commercial organic egg production. Animal Welfare ,12, 649-653.

Ito K, Weary DM and von Keyserlingk MAG, 2009. Lying behavior: assessing within- and between-herd variation in free-stall-housed dairy cows. Journal of Dairy Science, 92, 4412–4420.

Laboratory Animal Science Association (LASA), 2008. Final report of a LASA/APC Working Group to examine the feasibility of reporting data on the severity of scientific procedures on animals. Available from:

http://www.rspca.org.uk/servlet/BlobServer?blobtable=RSPCABlob&blobcol=urlblob&blobkey=id&blobwhere=1147695595721&blobheader=application/pdf

Main DCJ, Barker ZE, Leach KA, Bell NJ, Whay HR and Browne WJ, 2010. Sampling strategies for monitoring lameness in dairy cattle. Journal of dairy science, 5, 1970-1978.

OIE (World Organisation for Animal Health), 2009. SOP for OIE Validation and Certification of Diagnostic Assays. Available from http://www.oie.int/fileadmin/Home/eng/Our_scientific_expertise/docs/pdf/en_fichier_SOP.pdf

OIE (World Organisation for Animal Health), 2011. Terrestrial Animal Health Code. Available from http://www.oie.int/international-standard-setting/terrestrial-code/access-online/

Presi P and Reist M, 2011. Review of methodologies applicable to the validation of animal based indicators of welfare. Available from http://www.efsa.europa.eu/en/supporting/pub/171e.htm

Rushen J, Butterworth A and Swanson JC, 2011. Animal behavior and well-being symposium. Farm animal welfare assurance: science and application. Journal of Animal Science, 89, 1219–122.

Salman MD (Ed.), 2003. Animal Disease Surveillance and Survey Systems: Methods and Applications. Iowa State Press, Ames, 222 pp.

Smulders FJM and Algers B, 2009. Welfare of production animals: assessment and management of risks. Food safety assurance and veterinary public health, Volume 5, Wageningen Academic Publishers, Wageningen, The Netherlands, 588 pp.

Wathes CM, Kristensen HH, Aerts JM and Berckmans D, 2008. Is precision livestock farming an engineer's daydream or nightmare, an animal's friend or foe, and a farmer's panacea or pitfall? Computers and Electronics in Agriculture, 64, 2-10.

Welfare Quality® Protocol, 2009a. Welfare Quality® Assessment Protocol for cattle. Welfare Quality® 141 Consortium, Lelystad, The Netherlands, 180 pp.

Welfare Quality® Protocol, 2009b. Welfare Quality® Assessment Protocol for pig (sows and piglets, growing and finishing pigs). Welfare Quality® Consortium, Lelystad, The Netherlands, 122 pp.

Welfare Quality® Protocol, 2009c. Welfare Quality® Assessment Protocol for poultry (broilers, laying hens). Welfare Quality® Consortium, Lelystad, The Netherlands, 114 pp.

GLOSSARY

Accuracy (Validity): the overall correctness of an animal-based measure/welfare outcome indicator in identifying a welfare outcome.

Animal-based measure: a response of an animal or an effect on an animal used to assess its welfare. It can be taken directly on the animal or indirectly and includes the use of animal records. It can result from a specific event, e.g. an injury, or be the cumulative outcome of many days, weeks or months, e.g. body condition.

Composite animal-based measure: a new measure for an individual generated by a combination of several different animal-based measures from that particular individual.

Factor: any aspect of the environment of the animal, in relation to housing and management, genetic selection of animals, transport and slaughter, which may have the potential to improve or worsen the welfare of the animal.

Group-level animal-based measure: a measure that represents the mean, median or variation of the same animal-based measure from all individuals, or a representative sample of individuals, in a group of animals such as a pen, flock, herd or slaughter batch.

Hazard (in this context): a factor with the potential to cause poor welfare.

Intensity: the strength of a consequence.

Magnitude: a function of the intensity and duration of a positive or negative consequence on welfare.

Management-based measure: an evaluation of what the animal unit manager or stockperson does and which management processes or tools are used.

Measure: a form of evaluation rather than an intervention intended to deal with a problem.

Measurement: the result of the above evaluation, e.g. size and depth of wound, percentage of lame animals.

Non-animal-based measure: an evaluation of a factor of combination of factors (resources or management) that may be linked to change in the likelihood of good or poor welfare.

Predictive welfare indicator: an observation, a record or a measurement that does not indicate a substantial current welfare problem, but identifies an animal whose welfare is changing.

Quantitative risk assessment: a risk assessment that provides numerical expressions of risk and indication of the attendant uncertainties.

Reliability: a general term referring to the ability of the animal-based measures/welfare outcome indicator to be applied under various conditions, by different personnel while still providing similar (correct) results.

Repeatability: level of agreement between repeated "measurements" of the animal-based measure/welfare outcome indicator on the same "sample" by the same assessor, on different occasions.

Reproducibility: degree of agreement between measurements or observations conducted in replicates by different people.

Resource-based measure: an evaluation of a feature of the environment in which the animal is kept or to which it is exposed.

Risk assessment: a scientifically based process consisting of the following steps: (i) exposure assessment; (ii) consequence characterisation; and (iii) risk characterisation.

Robustness: the extent to which an animal-based measure/welfare outcome indicator is affected by changes in things such as environment or time of day.

Sensitivity: the minimum level of welfare outcome (changes) that will be detected by the animal-based measure.

Specificity: the extent to which an animal-based measure/welfare outcome indicator is specific for one welfare outcome, or relates (respond to) several outcomes.

Threshold: a cut-off value when a animal-based measure or welfare outcome indicator is considered to be "positive" i.e. indicative of a defined welfare outcome.

Validity: the fitness of an animal-based measure or a welfare outcome indicator that has been properly developed, optimised and standardised for an intended purpose. Validation includes estimates of the analytical and diagnostic performance characteristics of the measure/indicator (i.e. sensitivity and specificity).

Welfare indicator: an observation, a record or a measurement used to obtain information on an animal's welfare. This is usually measured but may not be but just show a trend.

Welfare outcome: a consequence for the welfare of an individual or group of animals of genetic selection or modification or of a period of housing, management, handling, transport, stunning or other treatment.

Welfare outcome indicator: an observation, a record or a measurement used to obtain information on an individual animal's welfare that can be reliably used in practice by trained people. It may be the outcome of genetic selection or modification or of a period of housing, management, handling, transport, stunning or other treatment.