Newcastle University e-prints

Robinson Library, University of Newcastle upon Tyne, Newcastle upon Tyne. NE1 7RU. Tel. 0191 222 6000

animal

Synthesis of semantic modelling and risk analysis methodology applied to animal welfare

M. B. M. Bracke[1†], S. A. Edwards[2], J. H. M. Metz[3], J. P. T. M. Noordhuizen[4,5] and B. Algers[6]

[1]Animal Sciences Group, Wageningen University and Research Centre, P.O. Box 65, 8200 AB Lelystad, The Netherlands; [2]University of Newcastle, School of Agriculture, Food and Rural Development, King George VI Building, Newcastle upon Tyne, NE1 7RU, UK; [3]Department of Agrotechnology and Food Sciences, Farm Technology Group, Wageningen University and Research Centre, P.O. Box 6700 AA, Wageningen, The Netherlands; [4]Department of Animal Health & Public Health, National Veterinary School of Nantes, Nantes, France; [5]Department of Obstetrics, Reproduction & Herd Health, Faculty of Veterinary Medicine, University of Gent, Belgium; [6]Department of Animal Environment and Health, Faculty of Veterinary Medicine, Swedish University of Agricultural Sciences, P.O. Box 234, SE-53223 Skara, Sweden

Decision-making on animal welfare issues requires a synthesis of information. For the assessment of farm animal welfare based on scientific information collected in a database, a methodology called 'semantic modelling' has been developed. To date, however, this methodology has not been generally applied. Recently, a qualitative Risk Assessment approach has been published by the European Food Safety Authority (EFSA) for the first time, concerning the welfare of intensively reared calves. This paper reports on a critical analysis of this Risk Assessment (RA) approach from a semantic-modelling (SM) perspective, emphasizing the importance of several seemingly self-evident principles, including the definition of concepts, application of explicit methodological procedures and specification of how underlying values and scientific information lead to the RA output. In addition, the need to include positive aspects of welfare and overall welfare assessments are emphasized. The analysis shows that the RA approach for animal welfare could benefit from SM methodology to support transparent and science-based decision-making.

Introduction

Over a number of decades, applied ethology has generated knowledge about animal behaviour, stress physiology, health and animal production to help answer empirical questions about animal welfare *aspects* of housing systems (e.g. whether pigs need straw; for extensive reviews see, e.g. SVC, 1995, 1996 and 1997). However, for sound ethical and political decision-making on animal welfare the information must be integrated further. To this end, new approaches are being investigated, sometimes drawing on other disciplines such as multi-criteria decision-making (Botreau *et al.*, 2007), social sciences (Beyer, 1998; Wemelsfelder *et al.*, 2001), HACCP (Hazard Analysis Critical Control Points) principles (Von Borell *et al.*, 2001) and multivariate analysis (Spoolder *et al.*, 2003a). Semantic modelling (SM) and Risk Assessment (RA) are potentially powerful tools in this new area of integrated animal welfare research.

For several years, the need to develop a Risk Analysis protocol for animal welfare has been recognized at the

European level (EC, 2000; EFSA, 2005). Recently, a report was published on qualitative RA applied to the welfare of intensively farmed calves (EFSA, 2006a and 2006b). This report made an essential step forward in supporting transparent decision-making on animal welfare. The new methodology, however, was recognized as being in need of further modification (CAC, 2002; EFSA, 2006a, p 8), and, because it will be used at the European level (European Food Safety Authority, EFSA) and may in the future even play a role worldwide (WTO – non trade concerns), it is important to examine and consolidate its scientific basis.

Over a number of years, work has also been carried out on so-called SM of animal welfare (Anonymous, 2001; Bracke *et al.*, 2002a and 2002b; De Mol *et al.*, 2004 and 2006). SM has been designed for the purpose of formalized assessment of animal welfare based on available scientific information, including scientific knowledge and scientific descriptions of housing systems in terms of (a combination of) both environment-based and animal-based measures (Bracke, 2007a; Bracke, 2008). It was originally developed for assessment of housing systems for dry sows (SOWEL model, see Bracke, 2001; Bracke *et al.*, 2002a and 2002b),

but it has also been applied to poultry (FOWEL, De Mol *et al.*, 2004 and 2006), to tail biting in pigs (PIGTAIL, Bracke *et al.*, 2004a and 2004b) and to assess enrichment materials for pigs (RICHPIG, Bracke *et al.*, 2007a and 2007b; Bracke, 2008). To date, however, while SM received some recognition (e.g. Rushen, 2003), it is far from being broadly implemented as a useful methodology for integrated welfare assessment (Blokhuis *et al.*, 2003; Spoolder *et al.*, 2003a and 2003b; Keeling, 2005; Aerts *et al.*, 2006; EFSA, 2006a and 2006b; Botreau *et al.*, 2007; Bracke, 2007a). A description of SM principles in relation to RA could illustrate the value for science-based, integrated welfare assessment.

Given the need to develop RA methodology for animal welfare further and the potential of SM, the objectives of this study were to evaluate the first formal RA applied to the welfare of calves (EFSA, 2006a and 2006b) from an SM perspective and to formulate suggestions for its improvement.

Methodologies

This section introduces the subjects of SM, RA generally, and the RA approach applied to calf welfare in the recent EFSA (2006a and 2006b) report as the starting point for the analysis.

Semantic modelling

SM aims at a quantified assessment of (overall) animal welfare, based on a systematic, formalized review and analysis of all available scientific information (Bracke *et al.*, 2006). The meaning of words plays a crucial role in the translation of scientific information into welfare scores, hence the term 'semantic'. The ultimate objective of SM is to support transparent, science-based decision-making, but its more proximate objective is to establish quantified welfare assessment as a scientific discipline, comparable to 'evidence-based medicine'. To this end, underlying values have been made explicit, and sharp distinctions have been made between the descriptive, normative and ethical aspects of animal welfare. SM is a descriptive activity (i.e. dealing with matters of 'fact'), assessing the level of welfare of the animals on a farm or in a pen as accurately as possible from a biological point of view, performed within a detailed set of normative methodological rules. All other aspects of welfare (e.g. ethical and political decisions, and deciding what levels of animal welfare are acceptable) are considered to be outside its present scope.

In SM, animal welfare has been defined as the quality of life as perceived by the animals themselves (Bracke *et al.*, 1999a). Welfare is a function of the animal's needs, which relate to the animal's biological control systems that have developed in the course of evolution to deal with a variable environment (Wiepkema, 1987). Biological needs include the need for food, water, rest, social contact, reproduction-related needs (such as the need for a mate and maternal behaviour), movement, exploration (including foraging and play), body care, elimination (voiding of faeces and urine), thermoregulation, respiration, health and safety (Bracke *et al.*,
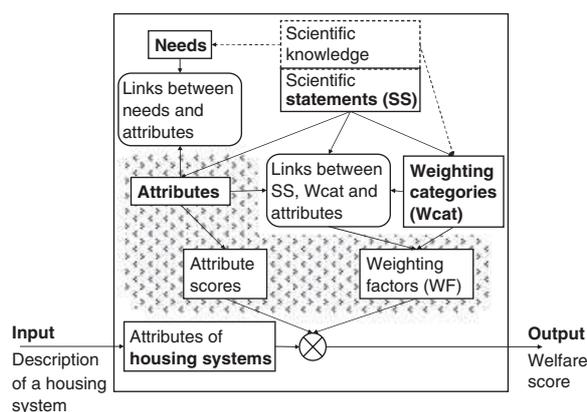
Figure 1 Structure of a semantic model for animal welfare assessment implemented in a relational database with linked tables. Primary tables are shown in bold. The dotted area covers the welfare model. Figure modified after Bracke *et al.* (2002a).

1999b and 1999c; Anonymous, 2001). These needs can, somewhat loosely, be categorized as behavioural and physiological needs (Bracke *et al.*, 1999c). All needs, however, have not only an environment-based aspect (food, water, space) but also a behavioural dimension (e.g. searching for food and consuming food), a (stress- and patho-) physiological dimension (e.g. fat deposition, emaciation, disease) and, most important for welfare, also an emotional dimension (e.g. feeling hungry). To assess the overall welfare, all needs are to be assessed with respect to the dimensions of intensity, duration and incidence (Willeberg, 1991; Anonymous, 2001) of both welfare performance criteria and underlying welfare-relevant emotions.

A semantic model takes a description of a housing system as input and produces a weighted welfare score on a scale from 0 to 10 as output (see Figure 1). Attributes describe the welfare-relevant properties of housing systems (Bracke *et al.*, 2002a). The attributes have also been called 'assessment criteria' (Bracke *et al.*, 2007b; Bracke, 2008). Attributes refer to the animal's resources, such as food, space and social conditions. Attributes include (information about) both the so-called design criteria (e.g. food provision) and welfare performance criteria (e.g. body condition and health status), which correspond largely to environment-based and animal-based parameters, respectively (Anonymous, 2001). The attributes and the attribute scores (which are scores for each of the attribute's levels, e.g. 'severely restricted feeding', 0; 'restricted feeding', 5; and '*ad libitum* feeding', 10) are derived from scientific statements (see below).

The core activity in SM concerns the analysis of scientific statements, which typically describe if–then relationships between design and performance criteria (Anonymous, 2001; Bracke *et al.*, 2002a; Bracke, 2007a; Bracke, 2008). For example, when (if) deprived of food, (then) an animal will show increased foraging behaviour and reducing body weight. From the analysis, attributes are identified. This could be called 'attribute identification' by analogy to Hazard Identification in RA (see below). In addition, attribute scores

are assigned following the science-based ranking of its levels. For example, the different levels of 'food' could be different amounts of food ranked from worst to best based on scientific studies showing effects on animal welfare in the short and long term. Finally, weighting factors (WF) are calculated using so-called weighting categories, which give a classification of welfare performance criteria, roughly matching the scientific paradigms to measure (aspects of) animal welfare. These include (the study of) animal pain, health and disease, productivity, survival and fitness, stress physiology, frustration and avoidance, abnormal behaviour, natural behaviour, preferences and demand (Bracke *et al.*, 2002a). All paradigms are relevant for the assessment of overall welfare, including its negative and positive aspects. The studies of natural behaviour, preferences and demand are the main sources of information about positive welfare.

WF are calculated as the differences between the most negative welfare information available about an attribute's worst level and the most positive information about its best level (see Bracke *et al.*, 2002a). Formally, therefore, the importance (WF) of an attribute can only be determined when its range is known. The attribute 'food', for example, is much more important when the level 'no food at all' is included, because animals will die from lack of food. Food is much less important when all animals in the assessment domain receive largely adequate nutrition, as is normally the case in modern livestock production.

Welfare scores are calculated as weighted-average attribute scores. To allow for this simple, most parsimonious calculation rule, a number of principles have been formulated. Firstly, the attributes must be relevant from the animal's point of view. For example, food is relevant, because an animal can feel hungry. Secondly, attributes must be defined as much as possible as mutually independent welfare components ('sub-needs') and must apply to all housing systems in the assessment domain. The assessment domain contains all housing systems conceivable within the modelling assignment. Each housing system is defined in SM as a conglomerate of welfare-relevant attributes, and is represented only once in the assessment domain. Thus, whereas in the real world some systems are more prevalent than other systems, this distribution has not been relevant for SM studies published so far, because in these studies welfare has been assessed at the housing system level (and it would require presently largely lacking information about population attributes to derive population estimates for animal welfare). Finally, together the attributes in the model must cover welfare overall, i.e. all welfare needs and all possible housing systems in the assessment domain, while a representative sample of systems, covering the entire welfare range, is described and assessed in the 'model', in order to provide benchmarks giving direction to the user and reducing assessment subjectivity (Bracke *et al.*, 2002a; Bracke, 2008).

The rather elaborate modelling structure in SM is designed to generate the best possible welfare assessment based on 'all' available information by linking scientific information systematically to the information about the animals and their living conditions on a farm or in a pen/enclosure. As part of SM research, validation studies have been performed using comparisons of model outcomes with expert opinion (e.g. Bracke *et al.*, 2002b; Bracke *et al.*, 2007a and 2007b), (sensitivity) analysis (Bracke *et al.*, 2002a; Bracke, 2008) and empirical validation studies (Bracke *et al.*, 2004b; Bracke, 2007b). In these studies, the semantic models performed well, as indicated, for example, by high Spearman's correlation coefficients for overall welfare scores given by experts compared to model scores ($\rho > 0.9$). Despite this, an ongoing need for further development is recognized, for example, towards more empirically validated, probabilistic and participatory welfare assessment. Because an ongoing need for further development was foreseen in SM, available information is stored and processed in linked tables in a relational database in order to facilitate upgrading when new information becomes available.

Risk Assessment in general

RA is an established scientific activity in the area of human and animal health (e.g. NRC, 1983; CAC, 1999, 2002; EC, 2000, 2003; Health Canada, 2000; OIE, 2004a and 2004b; EFSA, 2005). The main objective is to support prioritizing available resources for Risk Managers. Whereas 'hard' quantitative RA is the ideal, RA is often more qualitative, because of limitations of the available scientific information. RA includes four interdependent activities, namely Hazard Identification (HI), Hazard Characterization (HC), Exposure Assessment (EA) and Risk Characterization (RC). RA may be defined as the 'evaluation of the likelihood of an adverse effect(s)' (EC, 2000, p 11), where risk is a function of the probability and severity of the adverse effect due to exposure to a hazard.

HI is defined as the activity of selecting so-called hazards, where a hazard is defined as an agent or situation having the potential to cause an adverse effect. Typical hazards are pathogens and toxins. Hazards can be identified for disorders in the domains of public health, food safety, animal health and animal welfare. In RA, hazards are commonly regarded as dichotomous (yes/no) variables. EA expresses the probability that an individual, human or animal, in a defined target population will be exposed to the hazard. EA is the probabilistic equivalent at the population level of the more deterministic attribute scores in SM describing welfare-relevant properties at the farm (housing system) level. HC, by contrast, is an assessment at the individual level, at least when applied to animal welfare. HC is defined as the compound assessment of both severity of the adverse effect and the probability of that effect occurring as a consequence of exposure to the hazard. RC is a weighted score for risk related to a hazard, taking into account both its EA and HC.

RA is part of Risk Analysis. Other components of Risk Analysis, which will not be discussed here in more detail, are Risk Management and Risk Communication. Similar to SM, RA is conducted with a view to supporting transparent

Figure 2 Example of the main Risk Assessment Table as presented in the annexes of EFSA reports (2006a and 2006b) called 'Risk Characterization Scores'. This example shows the first five hazards reported for the housing system: 'White veal in small groups, not suckling'. HC: Hazard Characterization; EA: Exposure Assessment; RC: Risk Characterization.

and science-(evidence-)based Risk Management to be conducted by political decision-makers, and Risk Communication with affected or interested stakeholders. A first application of the RA principles to the field of animal welfare is described in the next section.

Risk Assessment on calf welfare
Recently, the report on calf welfare (EFSA, 2006b) was written by a group of seven scientists with expertise in animal science, ethology, veterinary medicine, RA and food safety (EFSA, 2006b, p. 69). The report presents an update of the previous SVC Report (1995) with an RA perspective, following the lines for RA of CAC (2002), but newly applied to the field of animal welfare. An opinion based on the report was also adopted by the Animal Health and Welfare Panel comprising 21 members (EFSA, 2006a). In this paper, a reference will often be made to both reports (EFSA, 2006a and 2006b), especially for reference to the RC table, which was an annex to each report (see Figure 2). Methodology of the approach was described rather broadly in the scientific report (EFSA, 2006b) and will be described in more detail by Algers *et al.* (in preparation). The main points are briefly presented here.

Because hazards were identified as factors that may compromise welfare needs, the first step was to describe the needs of calves. Next, in total 43 hazards were identified, which were assessed for nine housing systems with respect to HC and EA, each on a 5-point scale. Housing systems included, for example 'White veal in small groups, bucket fed (i.e. not suckling)', 'Hutches outside with replacement dairy calves, bucket fed (not suckling) + solid foods, weaned at 2 to 3 months', 'Feed lots (high-density groups within outside pen)' and 'Groups of dairy calves with an automatic feeding system (not suckling) + solid foods, weaned at 2 to 3 months'.

EA was defined as 'the quantitative and qualitative evaluation of the likelihood of hazards to welfare occurring in a given calf population' (EFSA, 2006b, p. 11, based on CAC, 2002), and this was operationalized as the perceived frequency of hazard occurrence in each housing system, with every 20% increase in frequency resulting in 1-point increase on the qualitative (subjective) 5-point scale. To confirm EA scores, a group of seven practicing veterinarians was consulted. They had extensive clinical experience in

calf medicine, covering calf production in the EU, and were not affiliated with the calf production industry (EFSA, 2006b, p. 70).

EA scores were given for all hazards considered relevant, except for six hazards where the label 'data not available' was given. These related to iron deficiency, allergic proteins, too rich diet (overfeeding), insufficient floor space, inadequate health monitoring and inadequate haemoglobin (Hb) monitoring.

HC was defined as 'the qualitative and quantitative evaluation of the nature of the adverse effects associated with the hazard' (EFSA, 2006b, p. 11). HC was also scored on a 1 to 5 scale.

When available, the HC and EA scores were multiplied for RC (scale 1 to 25). RC scores of 20 points or more were classified as 'major risk', and scores below 9 points were classified as 'negligible risk'. All intermediate scores were labelled as 'minor risk'.

No measure of uncertainty of the scores was reported, and no structured arguments were given relating the scientific facts in the review to the HC, EA or RC scores.

Comparison of RA and SM concepts
Both SM and RA aim to support transparent and science-based, ethical and political decision-making. A superficial comparison may suggest three main differences in the objective: SM aims to develop a scientific methodology for welfare assessment, whereas RA more pragmatically aims to support political decision-makers; SM is directed at overall scores, whereas RA focuses on components (hazards); and, whereas RA focuses on negative welfare, SM explicitly includes aspects of positive welfare too. Table 1 gives a more detailed account of the differences between the methodologies.

Underlying the differences, however, is a structural similarity, which (tentatively) equates hazards in RA and attributes in SM (see Figure 3). The similarity, which is also evident in Table 1 as the list of aspects used to identify differences, follows from the shared underlying objective of integration of scientific information to support decision-making. It provides the basis for the analysis of the EFSA (2006a and 2006b) reports from the perspective of SM as described in the remainder of this paper (and in Bracke *et al.*, submitted). In particular, with respect to the three main differences identified above, this paper will emphasize the need for scientific methodology, for overall assessments (even if the primary objective is 'only' to prioritize component hazards) and for inclusion of positive welfare aspects in the further development of RA applied to animal welfare.

Analysis of the EFSA calf welfare report
The analysis of the calf welfare reports (EFSA, 2006a and 2006b) from the perspective of SM included the following activities. Firstly, the reports (EFSA, 2006a and 2006b; SVC, 1995) were screened to identify points for discussion. Secondly, the welfare needs, hazards and their scores (HC,

Table 1 *Overview of differences between risk assessment and semantic modelling*

Aspect	Risk assessment	Semantic modelling
Prime objective	Support Risk Manager to avoid welfare problems	Establish quantified overall welfare assessment as scientific discipline to support ethical and political decision-marking
Scientific methodology	Statistics/mathematics	Formal logic (if–then) – system with procedures, principles and definitions
Scientific content	Veterinary/medical epidemiology	Animal behaviour, veterinary medicine, (stress-)physiology, animal science
Assessment framework	–/Needs	Biological conceptual framework and welfare needs (Anonymous, 2001)
Information processing	Literature review	Formalized analysis of systematically collected scientific statements
Assessment domain	Actual population of farms (e.g. in EU)	Housing systems defined by their welfare relevant properties
Factors	Hazards addressing negative welfare	Attributes addressing negative and positive aspects of welfare
Factor scale	Dichotomous (yes/no) hazards	(Semi-) continuous (welfare-relevant) attributes
Factor coverage	– (applied when considered relevant)	Common denominators of housing systems in the assessment domain
Factor requirements	Causal relation to a (welfare) problem	Welfare-relevant, i.e. matter to the animal's point of view
Factor list requirement	–	Cover overall welfare, i.e. all needs; minimized overlap to avoid double-counting
Occurrence score	EA score is a probabilistic concept	Attribute score is a truth-value, i.e. property of a housing system
Occurrence scale	EA is expressed on a 1 to 5 scale	Attribute scores range from 0 (worst level) to 10 (best level)
Factor importance	HC based on expert opinion (in EFSA, 2006a and 2006b)	Weighting factors (WFs) based on classified scientific evidence
Importance scale	HC is expressed on 1 to 5 scale	WFs have a relative scale reflecting the amount of differential scientific evidence
Component factor calculation	Multiplicative: RC = EA * HC	Weighted attribute score = attribute score * normalized WF
Overall impact	Overall risk is not assessed (in EFSA, 2006a and 2006b)	Overall welfare is expressed on a scale from 0 (worst) to 10 (best)
Calculation of overall impact	–	Overall welfare = sum of weighted attribute scores
Uncertainty	– (Not stated in EFSA, 2006a and 2006b)	Variability of expert opinion and uncertainty algorithm
Validation	– (Not in EFSA, 2006a and 2006b, but see Bracke *et al.*, submitted)	Expert opinion, (sensitivity) analyses and preliminary empirical studies

HC = hazard characterization; EA = exposure assessment; WF = weighting factor.
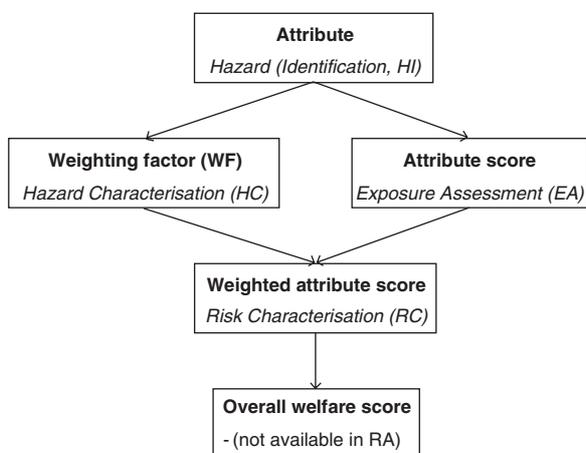
Figure 3 Terminology and abbreviations used in semantic modelling (SM, in bold) and Risk Assessment (RA, in italics) emphasizing similarities between RA and SM. Note, however, that concepts within blocks are related, but not identical (see text and Table 1).

EA and RC) were analysed using several classifications, including the list of needs formulated in Bracke *et al.* (1999c). Finally, a questionnaire was sent to experts to verify points identified in the analysis. The results of this survey, however, are reported in a separate paper (Bracke *et al.*, submitted).

The first screening of the paper resulted in the following five points.

Firstly, major welfare hazards identified in other reviews on calf welfare (e.g. SVC, 1995; Fraser and Broom, 1997; Anonymous, 2001) were not highlighted as clearly as expected. A lack of adequate roughage for normal rumen development, lack of social contact and insufficient space were identified as the most important design criteria for veal calves in Anonymous (2001), prioritized in the order listed here. In addition, that paper identified anaemia (iron deficiency) as a major welfare performance criterion. Instead, the main risks identified in the EFSA report all, directly or indirectly, related to health problems: inadequate

colostrum intake, inadequate ventilation, pathogen exposure, restocking (no 'all in–all out') and mixing of calves from different sources. By contrast, 'insufficiently balanced solid food', which could be referring to the lack of proper roughage, was classified as a 'minor risk', with an HC score of only 3. The problems with anaemia and space were addressed in the hazards 'iron deficiency resulting in haemoglobin levels below 4.5 mmol/l', 'inadequate haemoglobin monitoring' and 'insufficient floor space allowance'. However, for the two most relevant veal calf systems, these hazards were labelled as 'exposure data not available', and their level of risk was not reported.

Secondly, some hazards such as 'insufficient light' received a much higher HC score, whereas other hazards such as 'access to a natural teat' received a much lower score than expected based on an SM perception of how scientific evidence 'loads' onto welfare-relevant attributes.

Thirdly, the definition of EA, in terms of the frequency of hazard occurrence, appeared to be problematic in light of the dimensions of intensity, duration and incidence recognized in SM.

Fourthly, the list of housing systems did not include poor welfare systems, such as the former system of individual housing of veal calves in crates (now banned in the EU) and the individual housing of young calves in baby boxes (0.6 to 0.8 m wide) during the first 6 to 8 weeks of life (see SVC, 1995; Anonymous, 2001; EFSA, 2006b). Nor did the report include the perhaps most positive welfare system of calves reared with their dams on (semi-extensive) pasture in a herd.

Fifthly, detailed specifications of hazards and housing systems were not presented. Only 9 out of 43 hazards (21%) were defined in a specific way allowing assessment without further interpretation. These were as follows: iron deficiency (Hb below 4.5 mmol/l), no access to a natural or artificial teat, no bedding, social isolation, continuous restocking (no all in–all out), lack of maternal care, mixing calves from different sources, 'castration/dehorning, no anaesthetics' and separation from the dam. Twenty-two hazards (i.e. 51%) used the qualitative terms 'inadequate', 'poor', 'inappropriately' or 'insufficient' in the description.

The second part of the analysis involved applying classifications. One classification showed that in total 22 hazards, i.e. 51%, primarily related to animal health. Much lower percentages were found for the other needs: food (9%), thermal comfort (9%), safety (7%), social contact (5%), exploration (5%), rest (5%), respiration (5%), kinesis (i.e. movement, 2%) and water (2%). Only one hazard related somewhat to the need to forage (i.e. to search for food). No hazards were found addressing the need to groom and play. A second classification showed that 37 hazards (86%) were mainly addressing (patho-) physiological needs, whereas only six hazards (14%) primarily addressed behavioural needs. Comparison of EA scores across housing systems identified 14 hazards which had no variance, i.e. no difference between the highest and lowest EA score given.

Discussion of the SM perspectives on RA methodology for animal welfare

Definitions

A first point for discussion concerns the operationalization of EA in terms of frequency of hazard occurrence. In SM, informed common-sense reasoning is applied to all welfare assessment claims. From that perspective, risk is a function of effect magnitude and its occurrence probability. In RA, effect magnitude is primarily covered by HC, whereas occurrence probability relates primarily to EA. When EA is defined in terms of hazard frequencies (i.e. 'incidences' or 'prevalences'), common-sense reasoning leads to the suggestion that hazard durations should also be included, either in EA or in HC, although not in both as this would lead to double-counting in RC. In microbiological RA, durations have been included in the definition of EA (CAC, 1999, p. 4), but the question remains whether this applies to animal welfare. In SM, the magnitude of a welfare problem is defined by the intensity, duration and incidence of the effect, not the cause. A cause (hazard/design criterion) may last for a long time, but if the effect as defined in HC is short, the welfare impact (risk) may be small. Only when the underlying science has shown that a *unit* of exposure results in a specified effect and when multiple effects are known to be additive may hazard frequencies be multiplied by HC to determine RC. In veterinary epidemiology, the effects are often not additive (Noordhuizen *et al.*, 2001; Thrusfield, 2005), and the relevant information is mostly lacking in animal welfare science today. Therefore, from an SM perspective three suggestions are made in relation to the definition of EA in the EFSA (2006b) report: firstly, to take account of durations; secondly, to ensure HC correctly relates probability and severity of effect to the duration of exposure; and, thirdly, to specify hazards in close relation to the underlying science, thus providing a first illustration of the need to specify definitions and information processing rules.

Logical analysis in SM has shown that science essentially describes if–then relationships between design criteria/hazards and welfare consequences (Anonymous, 2001; Bracke, 2007a; Bracke, 2008). This implies 3 rather than 2 primary scores for RA: the probability that 'if' will occur (1), the probability of the if–then relationship (2) and the severity of the 'then' consequences (3). The first primary score (probability that 'if') is the EA score, which, depending on the underlying science, takes into account possible variation of hazard duration and the frequency across time in an individual's life, across individuals on a farm and across farms. In SM, the latter component has not been incorporated so far, mainly because of a lack of scientific information on the prevalence of the various welfare hazards across (random samples of) farms and because its prime objective has been methodologically sound welfare assessment based on available scientific information (Bracke *et al.*, 2002a). The second and third scores are both incorporated in the HC score: the probability of the if–then

relationship as revealed by science, and the severity of the welfare effect as perceived by the animal.

Such a theoretical analysis of concepts is not merely a matter of (arbitrary) semantics. One implication could be that EA (probability that 'if') and the first component of HC (probability that 'if–then') should be assessed on a scale between 0 and 1, or between 0% and 100%, in accordance with the convention of expressing probabilities in statistics. Perhaps more important, the separation of these two functionally related components of occurrence probability (that if and that if–then), and the lumping of the latter with effect magnitude in HC have resulted in an RA classification that differs from the common-sense view of risk as a function of effect magnitude and (effect) occurrence probability. This difference originates in different perspectives: The RA conducted by EFSA is designed around causal hazards in order to support the Risk Manager in controlling risks, while the general public perceives risk primarily in terms of negative experiences (related to effects). This may have consequences for RA because cause–effect relationships are variable, with respect to both the number of causes related to an effect and the clarity of the relationship. For example, the consequences of the hazards 'exposure to pathogens' and 'inadequate colostrum intake' are much better understood than the consequences of the hazard 'insufficient light'. This may not only hamper subsequent Risk Communication but also compromise RA itself. For example, it may lead to accidental double-counting or underestimation of risk components when risk assessors would accidentally confuse the two modes of assessment. It could also lead to a relative overestimation of hazards with known adverse consequences (e.g. health-related hazards), and to an underestimation of effects (e.g. abnormal behaviours) with less well-understood causes (e.g. behavioural restrictions due to limited space).

Data not available

In the EFSA (2006a and 2006b) report, the use of the EA label 'data not available' implied that data were available for reported EA scores. Contrary to its suggestion, however, here the term 'data' does not refer to 'scientific data', because otherwise the report would have included references. The label 'data not available' meant that the authors and veterinary experts were not able to give subjective, experience-based EA scores.

Surprisingly, experts with 'extensive clinical experience in calf medicine, covering calf production in the EU' (EFSA, 2006b, p. 70), could not assess 'insufficient floor space allowance', but they were able to assess, for example, 'insufficient light'. Most noteworthy, however, was the label 'data not available' for the hazard 'iron deficiency, resulting in haemoglobin levels below 4.5 mmol/l', especially because the previous report had stated that a large standard deviation had been found in a large sample of commercially reared veal calves (Morisse et al., in press, cited in SVC, 1995) and that 'in many commercial white veal units it is routine practice to blood sample calves on several

occasions during growth and adjust iron intake so as to (just) meet the target haemoglobin concentration (Hb of 4.5 mmol/l) at slaughter' (SVC, 1995, p. 51). Therefore, some, but perhaps not the most recent, data may have been available to the calf report working group. As a consequence, the label 'data not available' may have suggested a general need for further research, and this may have masked the more specific need (of the Risk Manager) to consider addressing the industry's issues of data confidentiality. This suggests that a wider methodological framework is needed where other difficulties of the assessors, such as cost, resources, time and possible conflicts of interests, are identified and their possible consequences described (CAC, 1999; Beekman et al., 2007; Bracke et al., 2008).

Uncertainty

A further point for discussion concerns the need to include measures of uncertainty and variability. Several reports on the harmonization of RA (CAC, 1999; EC, 2000 and 2003; Health Canada, 2000; EFSA, 2005) recommended incorporation of measures of uncertainty based on a weighing of the scientific evidence. The EFSA (2006b) report, which reviewed the literature, did not include a numerical expression of uncertainty. As a result, reported EA, HC and RC scores may have given a false impression of accuracy and group consensus, which may be easily misinterpreted and be misused in the absence of information, which puts the scores into context (Health Canada, 2000).

Results from SM, where models have been 'validated' against expert opinion, have shown that complete consensus is rare (Bracke et al., 2002b, 2007a and 2007b; see also Bracke et al., submitted). In SM, the variation around the scores has been proposed as an indication of the confidence interval for model scores, which highly correlated with median expert scores (Bracke et al., 2002b). This was not sufficient, and an additional measure of uncertainty has been incorporated in the algorithm for calculating WF (Bracke, 2008). These are still rather crude measures of uncertainty, which, nevertheless, have been useful in formulating directions for empirical research (Bracke et al., 2004a; Bracke, 2007b) and could be used as starting points for improved dealing with uncertainty in SM and in RA for animal welfare.

Dichotomous hazards and scientific basis

In the EFSA (2006a and 2006b) report, all hazards were dichotomous (yes/no) variables. Only nine hazards were specifically defined, including one hazard for which a threshold value was specified: 'iron deficiency (Hb below 4.5 mmol/l)'. Most hazards, 51% out of the total of 43 hazards, were qualitatively defined, using terms like 'inadequate', 'poor' and 'insufficient', without specifying the level below which it was considered 'insufficient'. By contrast, the attributes in SM have a more continuous nature due to their close relationship to the underlying scientific information and because setting thresholds is

inherently arbitrary (Mendl, 1991). In the case of Hb deficiency, values (just) above 4.5 mmol/l have been shown to be associated with serious welfare problems, such as exercise intolerance (Piguet *et al.*, 1993) and stress (Van Reenen *et al.*, 1999). More importantly, for most hazards/attributes, when values are increased from just below a threshold to just above it, the improvement in welfare tends to be minimal. This is an important reason as to why political decision-makers are presently looking for alternative ways to improve welfare, for example, by issuing prescriptions of goals rather than means (see also Bracke, 2008; Bracke *et al.*, 2007a and 2007b). Finally, from an SM perspective (e.g. Bracke *et al.*, 2002a, 2007a and 2007b), the decision-maker/Risk Manager must set the thresholds, because, whereas SM restricts itself to the descriptive task of assessing the factual welfare status as best we can scientifically, the decision-maker must specify what level is acceptable/sufficient politically and ethically, taking into account other concerns such as economic, environmental and food safety considerations. These are formally separate tasks (e.g. CAC, 2002).

Although scientists may use terms like 'inadequate' and 'insufficient' as shorthand descriptors, they are best avoided, because they have prescriptive elements in their (semantic) meaning for non-scientists and because they may create a false sense of consensus, whereas the scientific interpretation may vary between, for example, what is insufficient in case of cruelty and neglect, and what is insufficient for animal welfare under normal farming conditions. Therefore, to support transparent decision-making, hazard levels must be specified in relation to the underlying science and practices. In order to explain how this is carried out in SM, consider the following quote:

'… dairy calves kept, from birth to 1 month of life in larger stalls (1.00 m × 1.50 m) showed a higher percentage of lying behaviour and grooming than calves kept in smaller stalls (0.73 m × 1.21 m). … and … lymphocyte proliferation was significantly higher in calves reared in large stalls (Ferrante *et al.*, 1998).' (EFSA, 2006b, p. 45).

This scientific statement provides support for distinguishing two levels for the attribute 'space', namely 'small stalls (0.73 × 1.21 m)' and 'large stalls (1.0 × 1.5 m)'. Other scientific studies, which must also be taken into account, will add other levels, resulting in a (semi-) continuous scale, with levels ranging from worst (small stalls) to best (e.g. large pen/pasture). In addition, the statements provide information about welfare consequences (here: changed lying, grooming and lymphocyte proliferation), which are related to the ranked levels and which allow calculating the WF of attributes.

Relationships between hazards

In SM, overall welfare is assessed as a weighted average of component attributes scores, where all available and relevant scientific information is taken into account. When RA would adopt this objective, the following requirements can be formulated with respect to 'Hazard Identification' (Bracke *et al.*, 2002a; Bracke, 2008): the list of hazards must be a manageable and consistent list of common denominators that apply across the assessment domain, for example, all calf housing systems in Europe, and the hazards should overlap as little as possible and together cover overall welfare (i.e. all welfare needs and both negative and positive aspects of welfare).

In order to illustrate the latter requirement, consider three of the most important hazards for white veal calves in small groups identified in the EFSA (2006b) report: 'inadequate colostrum intake – duration', 'mixing of calves from different sources' and 'exposure to pathogens'. In SM that is designed for assessing welfare and not for assessing pathogenetic disease mechanisms, each hazard/attribute must 'load' on welfare and, thus, matter to the animals from their point of view. Not one of the three main hazards under consideration is directly relevant from the animal's point of view. What is relevant is the resulting disease risk (posed, e.g. by feeding milk rather than colostrum), because it means that the animals will *feel* ill. In SM (Bracke *et al.*, 2002a), these hazards would not have qualified as 'attributes', but would have been identified as 'influencing factors', because they causally affect a welfare-relevant attribute ('health status'). The main problem with identifying influencing factors as welfare hazards/attributes is that their HC (and its equivalent WF calculation) may suffer from interactions. In the present example, the welfare impact of poor health due to 'exposure to pathogens' will depend heavily on 'inadequate colostrum intake' and 'mixing of calves from different sources', and may also depend on other factors such as transport stress and preventive medication.

As a second issue, the requirement of internal consistency implies that general principles are applied for HI. For example, in EFSA (2006a and 2006b), three separate hazards were formulated with respect to colostrum intake, namely inadequate quantity, quality and duration. This is problematic for two reasons. Firstly, it is more parsimonious to subsume the three dimensions under one hazard 'inadequate colostrum intake'. Secondly, when a differentiation is made, for example, between quantity and quality for 'colostrum intake' in order to assist the Risk Manager, then internal consistency requires making the distinction for all similar cases (such as space, contact with humans and bedding) where some rule is formulated as to what counts as 'similar cases'. In the EFSA (2006a and 2006b) report, as many as five different hazards were formulated related to 'poor floor conditions': gaps too large, too abrasive, too slippery, too dirty, wet floor for lying. As a result, 'poor floor conditions' may not have been identified as a major hazard, while it had previously been identified as a very important (the fourth most important) welfare design criterion (Anonymous, 2001). By splitting up some hazards in components and lumping other hazards in broad classes (e.g. the main hazard 'exposure to pathogens' could be divided according to type of pathogen and pathogen load), the RC in EFSA (2006a and 2006b) may have been distorted. Scores for hazard components may be provided to

explain hazard scores and to assist the Risk Manager. To avoid misinterpretation, however, it is important to use general and rational principles (good reasons) in a methodologically described set of procedures when formulating a list of hazards.

Scientific evidence for HC

In SM, scientific information from different sources is used to formulate hazards (HI, see above) and to derive WF, i.e. for HC. In the EFSA (2006a and 2006b) report, HC scores were not explicitly related to scientific information, and this may hamper scientific scrutiny and the search for improvements. For example, 'insufficient light' has invariably been given the highest possible HC score of 5, whereas relatively little evidence was cited in the reports (SVC, 1995; EFSA, 2006b). Without the data, however, the justification for the high HC score cannot be verified. In contrast, the hazard 'access to a natural teat' received an HC score of only 2, even though much evidence exists and is cited in the EFSA and SVC reports, that calves have a *strong* need for sucking on a teat. Calves kept with the cow (with a 'natural teat') will often refuse to drink milk from a bucket (EFSA, 2006a, p. 12) and 'nipple feeding greatly reduced the incidence of navel sucking' (SVC, 1995, p. 53). Thus, contrary to the HC scores reported in EFSA (2006a and 2006b), it can be argued that when all scientific evidence is taken into account, welfare may be reduced more when a calf is deprived of a natural teat than when deprived of light. Specification of the scientific evidence for HC scores would allow verification and falsification, thus providing a point of entry for science in RA.

Risk Characterization (RC)

In the EFSA (2006a and 2006b) report, RC scores were calculated by multiplying EA and HC scores. From a statistical point of view this may be correct as it involves two probabilities (that if and that if–then) and one measure of severity (of 'then'). Nevertheless, objections can be made against this procedure. The outcomes of statistical RA may differ considerably from common sense, possibly because the multiplication of discrete EA and HC scores results in non-continuous RC scores and because errors are also multiplied. For these reasons, SM has so far used the more simple and parsimonious weighted-average (i.e. additive) calculation rule, and this may have made models relatively robust (Bracke *et al.*, 2002b; Bracke, 2008).

An example of multiplication of error may concern the hazard 'poor floor conditions'. The five hazards related to 'floor conditions' received RC scores between 2 and 15. This is low, not only compared with the general view that hard and slippery wooden floors are a major problem for veal calves (Anonymous, 2001) but also compared with a hazard such as 'mixing of calves from different sources', which received the maximum possible RC score of 25. Floor conditions affect almost all veal calves all of the time, whereas the impact of mixing at the start of the rearing period (especially when managed with adequate colostrum

provision and preventive medication) is usually of a relatively short duration, and fewer calves within Europe are suffering from its adverse effects. From an SM perspective, it is therefore advised to use the three dimensions (intensity, duration and incidence) to express the magnitude of risk, and to use an additive rule for qualitative scores for the time being to avoid multiplication of error, even though in the end the multiplicative calculation rule may be preferred from a statistical/mathematical point of view.

A related point is that the Risk Manager's attention tends to be focused on the main risks. From an SM perspective, however, positive aspects of welfare, such as sucking a natural teat, must also be taken into account, and this requires a risk–benefit analysis where assessment of behavioural needs has been operationalized (Bracke and Hopster, 2006). In addition, Risk Managers may have to realize that whereas singular 'main' risks tend to be resilient to change, a group of minor risks may have a combined larger welfare effect, which is also more easily managed (Bracke *et al.*, 2004a). This draws focus of main risks towards the notion of overall welfare assessment.

Overall scores for housing systems

Comparing the multiplicative rule in RA with the additive rule in SM is not completely accurate, because RC scores are equivalent to 'weighted attribute scores' (see Figure 1), which are also derived by multiplication, namely from WF and attribute scores. This differs from RA, however, in that attribute scores express truth values, i.e. properties that do or do not apply to a given housing system.

The EFSA (2006a and 2006b) report mainly focused on environment-based, welfare-input hazards. From an SM perspective, the link between the input and output, i.e., respectively, environment-based design criteria and animal-based performance criteria, is very important (Anonymous, 2001; Bracke, 2007a). That was the reason, for example, that in Anonymous (2001) design and performance criteria were prioritized separately. Although the prioritization of the design criteria was based on the associated performance consequences, the prioritization of the performance criteria was based on the associated welfare impact in terms of intensity, duration and incidence. That study, in which 22 international experts were consulted anonymously, using the so-called Delphi method, may be regarded as a predecessor study of RA applied to animal welfare, where housing systems were also ranked for welfare based on their welfare-relevant properties (i.e. their design and performance criteria).

Unlike RA, SM is designed to derive overall welfare scores from attributes/hazards. However, whereas overall welfare is not made explicit in RA, it implicitly requires a notion of overall welfare, because HC involves assessing the severity of adverse effects and this is not possible unless some notion of overall welfare is presupposed. In SM, the impacts of the attributes on overall welfare are made explicit for a wide range of housing systems. In order to construct a balanced list of attributes/hazards, it is

important to ensure a proper 'test' population of housing systems, covering the whole logical space of possible welfare-relevant attributes (rather than what is actually present in reality). In particular, housing systems should be included defining the end-points of the overall welfare scale, because these may be regarded as a kind of positive and negative controls. In the EFSA (2006b) report, however, the list of housing systems did not include possibly most poor welfare systems, such as the former system of individual housing of veal calves in crates (now banned in the EU) and the individual housing of young calves in baby boxes (0.6 to 0.8 m wide) during the first 6 to 8 weeks of life (SVC, 1995; Anonymous, 2001). Nor did it include the perhaps most positive welfare system of calves reared with their dams on (semi-extensive) pasture in a herd. Although these systems were outside the remit of the mandate for the working group, leaving them out may have obscured the welfare progress made in the past and the possible progress in the future, i.e. the inclusion of positive and negative controls is important to help confirm the validity of the system by placing identified welfare problems in a broader perspective. Overall welfare assessments do not simply supply this additional information to the decision-maker/Risk Manager, making overall assessments explicit is also likely to benefit the component welfare prioritizations in RA.

Underlying values

Finally, the need to make underlying values explicit has been emphasized in the field of animal welfare (Sandøe and Simonsen, 1992; Fraser, 1995; Fraser *et al.*, 1997). In the analysis, a discrepancy was identified between the expected and reported hazard prioritization. Whereas iron deficiency, roughage, social contact and space were the expected major hazards (Anonymous, 2001), colostrum, ventilation, pathogen exposure, all in–all out and mixing of calves were reported to be major hazards in EFSA (2006a and 2006b). To explain this discrepancy, it may be noted that reported hazards were primarily health-related, whereas expected hazards were primarily behaviour-related. The difference may therefore, reflect different underlying perspectives or perceptions of animal welfare.

This suggestion was confirmed when hazards were classified: 86% addressed (patho-) physiological needs, and only 14% addressed behavioural needs. In addition, no hazard specifically addressed grooming or play behaviour. Elsewhere in the EFSA (2006a and 2006b) report, grooming was considered important (which is in line with it being one of the original five freedoms proposed in the Brambell committee (1965)), but play was not recognized as a separate need. This may be surprising, because play is considered important for all young animals, including calves (Fagen, 1981; Fraser and Broom, 1997; Anonymous, 2001; Spinka *et al.*, 2001). Whereas differences between behavioural and veterinary perspectives on welfare may have contributed (though this was not fully confirmed in a subsequent questionnaire survey, see Bracke *et al.*, submitted),

other, perhaps related, underlying values may also have played a role. The most important may be that the working group on calf welfare adopted an RA approach, not a risk–benefit assessment. As a consequence, positive aspects of welfare, such as natural behaviour and play, were not taken into account as welfare risks. Here, however, science's traditional focus on animal suffering may conflict with the more recent public-oriented perception of animal welfare emphasizing the fourth freedom, i.e. to express normal behaviour (FAWC, 1992; Fraser *et al.*, 1997; LNV, 2002; Bracke and Hopster, 2006). Underlying SM is a biological conceptual framework (Anonymous 2001), in which animals are the products of evolution where they have acquired the capacity to experience both negative and positive emotions (Fraser and Duncan, 1998; Spruijt *et al.*, 2001). Disease (absence of good health) is an important cause of negative emotions, and behaviour can be an important indicator of positive welfare.

Conclusions

The objectives of this study were to evaluate the EFSA (2006a and 2006b) report from an SM perspective, and to formulate suggestions for improvement for the RA methodology applied to animal welfare.

The EFSA report provided a first formal RA applied to animal welfare, thus providing an essential step towards transparent science-(evidence-)based decision-making at the European level. It followed established principles for (qualitative) RA developed in relation to human and animal health, and involved the steps of Hazard Identification (HI), Exposure Assessment (EA), Hazard Characterization (HC) and Risk Characterization (RC; CAC, 2002). SM was shown to be an independent, but closely related, approach for formal animal welfare assessment based on available scientific information. Both RA and SM are relatively new and potentially powerful approaches for integrated animal welfare assessment, which is a prerequisite for taking animal welfare into account in conjunction with other aspects of animal production (environmental impact, socio-economic aspects, product quality, human health and safety, etc.) in subsequent ethical and political decision-making.

The main strength of RA may be that it is suited to help Risk Managers prioritize resources in a negative, risk-averse coping strategy, as may apply to authorities and bodies whose role it is to develop regulations to avoid unnecessary suffering and disease. When the scope is widened to include positive welfare, as suggested by SM, risk–benefit analysis is called for. This may apply to decision-makers addressing the wider range of public concern, and, for example, to producer organizations and retailers developing welfare quality products. Most importantly, however, SM proposes that RA adopts (parts of) its more formal methodology for welfare assessment to further improve its performance from a scientific and decision-making point of view in the future.

References

Aerts S, Lips D, Spencer S, Decuypere E and de Tavernier J 2006. A new framework for the assessment of animal welfare: integrating existing knowledge from a practical ethics perspective. Journal of Agricultural and Environmental Ethics 19, 67–76.

Anonymous 2001. Scientists' assessment of the impact of housing and management on animal welfare. Journal of Applied Animal Welfare Science 4, 3–52.

Beekman V, de Bakker E and de Graaff R 2007. Standing on the shoulders of a giant: the promise of multi-criteria mapping as a decision-support framework in food ethics. In Sustainable Food Production and Ethics. Preprints of the 7th Congress of the European Society for Agricultural and Food Ethics (ed. W Zollitisch, C Winkcler, S Waiblinger and A Haslberger), pp. 95–100. Wageningen Academic Publishers, Wageningen.

Beyer S 1998. Konstruktion und Überprüfung eines Bewertungskonzeptes für pferdehaltende Betriebe unter dem Aspekt der Tiergerechtheit. PhD, Wissenschaflicher Fachverlag, Giessen.

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.

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

Bracke MBM 2001. Modelling of animal welfare: the development of a decision support system to assess the welfare status of pregnant sows. PhD, Wageningen University, Wageningen.

Bracke MBM 2007a. Animal-based parameters are no panacea for on-farm monitoring of animal welfare. Animal Welfare 15, 67–70.

Bracke MBM 2007b. Multifactorial testing of enrichment criteria: pigs 'demanded' hygiene and destructibility more than sound. Applied Animal Behaviour Science 107, 208–232.

Bracke MBM 2008. RICHPIG: a semantic model to assess enrichment materials for pigs. Animal Welfare (in press).

Bracke MBM and Hopster H 2006. Assessing the importance of natural behavior for animal welfare. Journal of Agricultural and Environmental Ethics 19, 77–89.

Bracke MBM, Spruijt BM and Metz JHM 1999a. Overall welfare assessment reviewed. Part 1: Is it possible? Netherlands Journal of Agricultural Science 47, 279–291.

Bracke MBM, Metz JHM and Spruijt BM 1999b. Overall welfare reviewed. Part 2: Assessment tables and schemes. Netherlands Journal of Agricultural Science 47, 293–305.

Bracke MBM, Spruijt BM and Metz JHM 1999c. Overall welfare reviewed. Part 3: Welfare assessment based on needs and supported by expert opinion. Netherlands Journal of Agricultural Science 47, 307–322.

Bracke MBM, Spruijt BM, Metz JHM and Schouten WGP 2002a. Decision support system for overall welfare assessment in pregnant sows. A: Model structure and weighting procedure. Journal of Animal Science 8, 1819–1834.

Bracke MBM, Metz JHM, Spruijt BM and Schouten WGP 2002b. Decision support system for overall welfare assessment in pregnant sows. B: Validation by expert opinion. Journal of Animal Science 8, 1835–1845.

Bracke MBM, Hulsegge B, Keeling L and Blokhuis HJ 2004a. Decision support system with semantic model to assess the risk of tail biting in pigs: 1. Modelling. Applied Animal Behaviour Science 87, 31–44.

Bracke MBM, Hulsegge B, Keeling L and Blokhuis HJ 2004b. Decision support system with semantic model to assess the risk of tail biting in pigs: 2. Validation. Applied Animal Behaviour Science 87, 45–54.

Bracke MBM, Zonderland JJ, Lenskens P, Schouten WGP, Vermeer H, Spoolder HAM, Hendriks HJM and Hopster H 2006. Formalised review of environmental enrichment for pigs in relation to political decision making. Applied Animal Behaviour Science 98, 165–182.

Bracke MBM, Zonderland JJ and Bleumer EJB 2007a. Expert judgement on enrichment materials for pigs validates preliminary RICHPIG Model. Applied Animal Behaviour Science 104, 1–13.

Bracke MBM, Zonderland JJ and Bleumer EJB 2007b. Expert consultation on weighting factors of criteria for assessing environmental enrichment materials for pigs. Applied Animal Behaviour Science 104, 14–23.

Bracke M, de Bakker E, Beekman V, Jansson K and de Graaff R 2008. TRAK: Voedselkwaliteit op het spoor – Op weg naar een transparant beleidsafwegingskader. (TRAK: on the track of food quality – towards a transparance policy decision framework), LEI report 2008-002, LEI, The Hague.

Bracke MBM, Edwards S, Metz JHM, Engel B, Buist WG and Algers B (Submitted). Expert opinion as 'validation' of Risk Assessment applied to calf welfare. Animal.

Brambell Committee 1965. Report of the technical committee to enquire into the welfare of animals kept under intensive livestock husbandry systems. Command Report 2836. Her Majesty's Stationary Office, London, UK.

CAC (Codex Alimentarius Commission) 1999. Principles and Guidelines for the Conduct of Microbiological Risk Assessment. CAC/GL-30, 7pp. http://www.codexalimentarius.net/download/standards/357/CXG_030e.pdf. Accessed 12 November 2006.

CAC (Codex Alimentarius Commission) 2002. Principles and guidelines for the conduct of microbiological risk assessment. Document CAC/GL 30.

De Mol RM, Schouten WGP, Evers E, Drost WC, Houwers HWJ and Smits AC 2004. Integrale welzijnsbeoordeling leghennen. (Integrated welfare assessment laying hens). Report 239. Agrotechnology & Food Innovations of Wageningen University and Research Centre, Wageningen.

De Mol RM, Schouten WGP, Evers E, Drost WC, Houwers HWJ and Smits AC 2006. A computer model for welfare assessment of husbandry systems for laying hens. Netherlands Journal of Agricultural Science 54, 57–168.

EC (European Commission) 2000. First report on the harmonisation of risk assessment procedures. Part 1: The report of the Scientific Steering Committee's Working Group on Harmonisation of Risk Assessment Procedures in the Scientific Committees advising the European Commission in the area of human and environmental health. European Commission, Brussles, Belgium.

EC (European Commission) 2003. Second report on harmonisation of risk assessment procedures by the Scientific Steering Committee (SSC). Appendix 3: Report on risk assessment of food borne bacterial pathogens: a quantitative methodology relevant for human exposure assessment. European Commission, Brussles, Belgium.

EFSA (European Food Safety Authority) 2005. Principles of risk assessment of food producing animals: Current and future approaches. Parma 2005, http://www.efsa.europa.eu/etc/medialib/efsa/science/colloquium_series/no4_animal_diseases/1179.Par.0017.File.dat/ses_summary_report_coll4_en1.pdf. Accessed 2 December 2006.

EFSA (European Food Safety Authority) 2006a. Scientific opinion on the risks of poor welfare in intensive calf farming systems. An update of the Scientific Veterinary Committee Report on the Welfare of Calves. EFSA-Q-2005-014. The EFSA Journal 366: 1-36. http://www.efsa.europa.eu/en/science/ahaw/ahaw_opinions/1516.html. Accessed 15 August 2006.

EFSA (European Food Safety Authority) 2006b. Scientific report on the risks of poor welfare in intensive calf farming systems. An update of the Scientific Veterinary Committee Report on the Welfare of Calves. EFSA-Q-2005-014 (144 pp), Annex to EFSA (2006a). http://www.efsa.europa.eu/en/science/ahaw/ahaw_opinions/1516.html. Accessed 15 August 2006.

Fagen R 1981. Animal play behaviour. Oxford University Press, New York.

FAWC (Farm Animal Welfare Council) 1992. FAWC updates the five freedoms. Veterinary Record 131, 357.

Ferrante V, Canali E, Mattiello S, Verga M, Sacerdote P and Panerai AE 1998. The effect of the size of individual crates on the behavioural and immune reactions of dairy calves. Journal of Animal Science 7, 29–36.

Fraser D 1995. Science, values and animal welfare: exploring the 'inextricable connection'. Animal Welfare 4, 103–117.

Fraser AF and Broom DM 1997. Farm Animal Behaviour and Welfare, 3rd edition. Baillière Tindall, London, UK.

Fraser D and Duncan IJH 1998. 'Pleasures', 'pains' and animal welfare: toward a natural history of affect. Animal Welfare 7, 383–396.

Fraser D, Weary DM, Pajor EA and Milligan BN 1997. A scientific conception of animal welfare that reflects ethical concerns. Animal Welfare 6, 187–205.

Health Canada 2000. Health Canada decision-making framework for identifying, assessing, and managing health risks. http://www.hc-sc.gc.ca/ahc-asc/alt_formats/hpfb-dgpsa/pdf/pubs/risk-risques_e.pdf. Assessed 2 December 2006.

Keeling LJ 2005. Healthy and happy: animal welfare as an integral part of sustainable agriculture. Ambio 34, 316–319.

LNV (Dutch Ministry of Agriculture, Nature and Food Quality) 2002. Beleidsnota Dierenwelzijn (maart 2002). LNV, The Hague, The Netherlands.

Mendl M 1991. Some problems with the concept of a cut-off point for determining when an animal's welfare is at risk. Applied Animal Behaviour Science 31, 139–146.

Morisse JP, Cotte JP and Huonnic D (In press) (as cited in SVC 1995). L'anémie chez le veau de boucherie. Evolution maîtrisée ou problème pathologique. Séminaire International sur le Veau de Boucherie.

Noordhuizen JPTM, Thrusfield MV, Frankena K and Graat EAM 2001. Applications of Quantitative Methods in Veterinary Epidemilogy. Wageningen Academic Publishers, Wageningen, The Netherlands.

NRC (National Research Council) 1983. Risk assessment in the federal government: Managing the process. US National Academies Press, Washington DC, USA.

OIE (Office International des Epizooties) 2004a. Handbook on Import Risk Analysis for Animals and Animal Products. Volume 1. Introduction and qualitative risk analysis. OIE, Paris, France.

OIE (Office International des Epizooties) 2004b. Handbook on Import Risk Analysis for Animals and Animal Products. Volume 2. Quantitative risk assessment. OIE, Paris, France.

Piguet R, Bruckmaier M and Blum JW 1993. Treadmill exercise of calves with different iron supply, husbandry, and work load. Journal of Veterinary Medicine Series A 40, 456–465.

Rushen J 2003. Changing concepts of farm animal welfare: bridging the gap between applied and basic research. Applied Animal Behaviour Science 81, 199–214.

Sandøe P and Simonsen HB 1992. Assessing animal welfare: where does science end and philosophy begin? Animal Welfare 1, 257–267.

Spinka M, Newberry RC and Bekoff M 2001. Mammalian play: training for the unexpected. Quarterly Review of Biology 76, 141–168.

Spoolder HAM, de Bree J, van der Mheen HW and Bracke MBM 2003a. Not animal related but housing parameters determine an expert's intuitive welfare assessment of pig welfare. In Proceedings of the 37th International Congress of the ISAE (ed. V Ferrante), p. 102. Fondazione Iniziative Zooprofilattiche e Zootecniche, Brescia, Italy.

Spoolder H, de Rosa G, Hörning B, Waiblinger S and Wemelsfelder F 2003b. Integrating parameters to assess on-farm welfare. Animal Welfare 12, 529–534.

Spruijt BM, van den Bos R and Pijlman FTA 2001. A concept of welfare based on reward evaluating mechanisms in the brain: anticipatory behaviour as an indicator for the state of reward systems. Applied Animal Behaviour Science 72, 145–171.

SVC (Scientific Veterinary Committee) 1995. Report on the welfare of calves. EC, Brussels, Belgium.

SVC (Scientific Veterinary Committee) 1996. Report on the welfare of laying hens. EC, Brussels, Belgium.

SVC (Scientific Veterinary Committee) 1997. The welfare of intensively kept pigs. Report to the Directorate General XXIV of the European Commission. Adopted 30 September 1997. Doc. XXIV/ScVc/0005/97. EC, Brussels, Belgium.

Thrusfield MV 2005. Veterinary Epidemilogy, 3rd edition. Blackwell Science Publishing, Oxford, UK.

Van Reenen CG, Mars HM, Leushuis IE, van Oirschot JT and Blokhuis HJ 1999. The effect of housing system and dietary iron supply on stress physiological measures and resistance against an experimental infection with bovine herpesvirus 1 (BHV1) in veal calves. Report. ID-Lelystad, Lelystad, The Netherlands.

Von Borell E, Bockisch FJ, 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.

Wemelsfelder F, Hunter TEA, Mendl MT and Lawrence AB 2001. Assessing the 'whole animal': a free choice profiling approach. Animal Behaviour 62, 209–220.

Wiepkema PR 1987. Behavioural aspects of stress. In Biology of Stress in Farm Animals: An Integrative Approach(ed. PR Wiepkema and PWM van Adrichem), pp. 113–133. Martinus Nijhoff, Dordrecht, The Netherlands.

Willeberg P 1991. Animal welfare studies: Epidemiological considerations. In Proceedings of the Society for Veterinary Epidemiology and Preventive Medicine (ed. MV Thrushfield), pp. 76–82. London.