The Old School, Brewhouse Hill, Wheathampstead,
Hertfordshire AL4 8AN, UK

Animal Welfare 2012, 21(S1): 73-79
doi: 10.7120/096272812X13345905673764
ISSN 0962-7286

Welfare risk assessment: the benefits and common pitfalls

LM Collins

School of Biological Sciences, Queen's University Belfast, Medical Biology Centre, 97 Lisburn Road, Belfast BT9 7BL, UK;
email: l.collins@qub.ac.uk

Abstract

Risk is defined as a situation involving exposure to danger. Risk assessment by nature characterises the probability of a negative event occurring and quantifies the consequences of such an event. Risk assessment is increasingly being used in the field of animal welfare as a means of drawing comparisons between multiple welfare problems within and between species and identifying those that should be prioritised by policy-makers, either because they affect a large proportion of the population or because they have particularly severe consequences for those affected. The assessment of risk is typically based on three fundamental factors: intensity of consequences, duration affected by consequences and prevalence. However, it has been recognised that these factors alone do not give a complete picture of a hazard and its associated consequences. Rather, to get a complete picture, it is important to also consider information about the hazard itself: probability of exposure to the hazard and duration of exposure to the hazard. The method has been applied to a variety of farmed species (eg poultry, dairy cows, farmed fish), investigating housing, husbandry and slaughter procedures, as well as companion animals, where it has been used to compare inherited defects in pedigree dogs and horses. To what extent can we trust current risk assessment methods to get the priorities straight? How should we interpret the results produced by such assessments? Here, the potential difficulties and pitfalls of the welfare risk assessment method will be discussed: (i) the assumption that welfare hazards are independent; (ii) the problem of quantifying the model parameters; and (iii) assessing and incorporating variability and uncertainty into welfare risk assessments.

Keywords: *animal welfare, expert opinion, policy, prevalence, risk assessment, welfare hazard*

Introduction

Risk is defined by the *Oxford English Dictionary* (OED 2010) as:

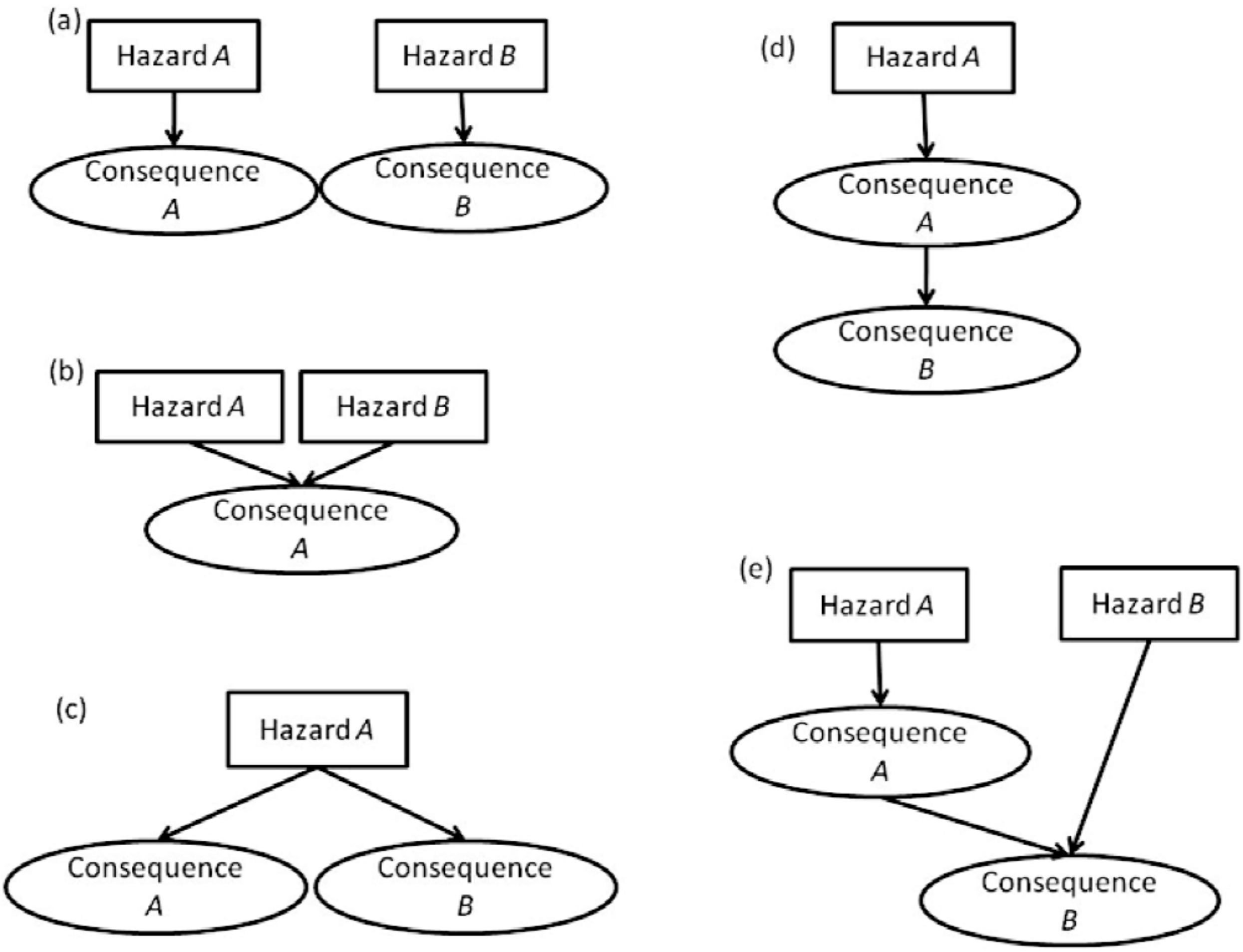
> Exposure to) the possibility of loss, injury, or other adverse or unwelcome circumstance; a chance or situation involving such a possibility.

Risk assessment characterises the probability of a negative event occurring and quantifies the consequences of such an event. The use of risk assessment methods is becoming increasingly common in the field of animal welfare. Ultimately, they provide a way of comparing the impact of very different welfare problems both at the individual and population level, within and between species, based on a number of key factors (EFSA 2008, 2009, 2010a,b; Collins *et al* 2010).

In risk assessment terminology, welfare problems are caused by a series of 'hazards'. By example, a brachycephalic head shape in dogs may be considered a potential welfare hazard with possible consequences, such as brachycephalic airway obstruction syndrome and a reduced ability to exercise (Asher *et al* 2009). Characterising a hazard, as the first step in the risk assessment process, is perhaps one of the most critical. The three most basic factors used to calculate risk are intensity of consequences, duration of effect of consequences (either as an absolute value if comparing within a breed or species, or as a proportion of lifetime if comparing between breeds or species, see Collins *et al* [2011], for an example) and prevalence (the proportion of affected individuals at any one time). These three factors allow a comparison of consequences and their impact on the animals experiencing them. However, in this basic form of risk assessment, specific details of the hazards are not considered.

To get a more complete picture of hazards and their consequences, and thus a more accurate risk estimate, it is important to consider information about the hazard itself in the calculation, for example by estimating the duration and probability of exposure to the hazard. In providing a quantitative, or even qualitative, value for each of these factors, the aim is to produce an objective estimate of risk for a series of potential welfare hazards.

Figure 1



Schematic diagram of the possible hazard and consequence associations. Although risk assessment assumes dependence between a hazard and its consequences, it assumes independence between hazards and the consequences of different hazards. This figure shows two hazards (boxes) and two consequences (ovals) and the potential range of associations: (a) an ideal scenario for risk assessment purposes, where Hazard A is independent of Hazard B and Consequence A is independent of Consequence B; (b) non-independence of hazards in that Consequence A can be caused by both Hazard A and Hazard B; (c) a hazard can have more than one consequence; (d) shows non-independence of consequences, further, that a consequence of one hazard (Consequence A) can sometimes be the cause of a secondary consequence (Consequence B), leading to an indirect link between Hazard A and Consequence B; (e) shows both non-independence of hazards and non-independence of consequences. Both Hazard A and Hazard B are associated with Consequence B, although Hazard A is associated indirectly through Consequence A.

Three key current pitfalls of welfare risk assessment procedures

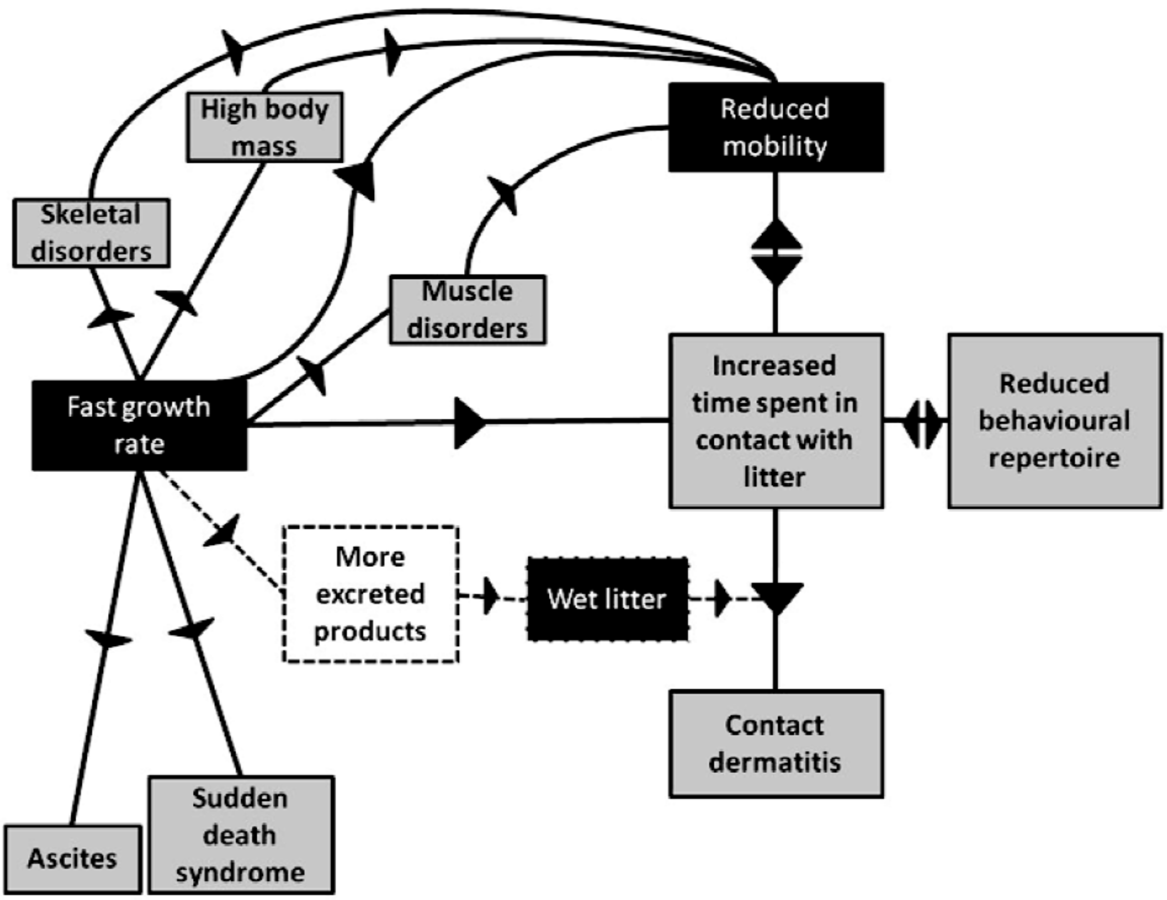
The assumption of independence

The assumption of independence is repeatedly made throughout risk assessment processes. First, there is the assumption that a hazard and its associated consequences are independent of other hazards and their associated consequences. Second is the assumption that intensity of consequences, duration of consequences and prevalence are independent. Third, in EO-based risk assessments, that the opinions of the individual experts are independent.

Whether the assumption that a hazard and its associated consequences and other hazards and their associated consequences are independent of each other is upheld in practice may be true for certain combinations of hazards (particularly those that are most dissimilar, with consequences affecting different, unrelated processes) (Figure 1[a]), but in many cases, there are varying degrees of overlap in the kinds of consequences caused by different hazards. In these cases, it is not always possible to know which of the potential hazards has led to the presentation of this effect, or whether the consequence is caused by the simultaneous actions of the two (or more) hazards (Figure 1[b]). The combination of different hazards and the effects of multiple hazards acting simultaneously on a population cannot be easily quantified. So it becomes even more complicated when multiple hazards can produce the same consequence, and that consequence, either independently or in unison

https://doi.org/10.7120/096272812X13345905673764 Published online by Cambridge University Press

Figure 2



The non-independence of hazards and consequences, using fast growth rate in commercial broilers as an example. Fast growth rate is shown to be linked both directly and indirectly to other hazards, which are considered as independent factors. Black boxes are hazards. Grey boxes are 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 adapted from Collins (2005).

with one of the hazards, leads to an increased likelihood of developing a second type of consequence (Figure 1[e]). For instance, in the commercial broiler chicken example given in Figure 2, the consequence 'increased time spent in contact with the litter' is associated with two hazards: reduced mobility, and fast growth rate. Are the two hazards acting together equally to give rise to the observed consequence, does one have a stronger association than the other, and can we consider the two hazards as independent when one is indirectly associated with the other (as fast growth rate is associated with skeletal disorders, high body mass and muscle disorders, all of which can lead to reduced mobility)?

The implications of this in terms of the results of the risk assessment are considerable. For those hazards that act in a non-independent manner with other hazards and consequences, the risk score may be substantially over- or underestimated, as the calculated scores in welfare risk assessments do not take into account second, or higher order consequences (ie consequences of consequences), and do not quantify the proportion of attributable risk associated with consequences caused by multiple hazards. In these cases, the prevalence data in particular are likely to be overestimates for each individual hazard as where reliable prevalence estimates have been found, they are unlikely to be known in sufficient detail to partition the risk appropriately between co-occurring hazards. This could ultimately alter the relative rank of the hazard risk scores. However, on the contrary, non-independence of hazards and consequences will not affect the scoring of the individual-focused factors of intensity of suffering and duration of suffering.

Although the number of different consequences for any single hazard may not affect the final aggregated scores of magnitude, welfare impact and risk score if the aggregated scores are standardised for the number of consequences, one might argue that if the actions of a hazard lead to multiple types of consequences, then standardising will mean that the risk outcomes for this hazard could be underestimated.

Non-independence is not an insurmountable issue, however, and in other areas of risk assessment, where the methods have been utilised for many years (such as engineering,

doi: 10.7120/096272812X13345905673764

financial markets and health and safety), associations between variables are incorporated into complex, multi-layer models predicting outcomes and optimising decisions based on quantified estimates of risk (Aven 2011). These can be developed using standard risk analysis packages such as @RISK (Palisade Corporation 2011), but do require estimates of at least some of the interaction outcomes so that likelihood distributions can be created. In those cases where this has been done, this has relied on large banks of data detailing previous co-occurrences of hazards with an outcome consequence (eg Anderson *et al* 2011; Qu *et al* 2011; Zabeo *et al* 2011).

Intensity and duration of consequences are considered and scored as independent factors in the risk assessment process. However, the assumption of independence here may once again be flawed as an individual's level of suffering is likely to be a function not only of the current intensity, but also how long it has been attempting to cope with the hazard it is experiencing, whether this has impacted on body condition and to what extent this has affected immune, endocrine and nervous-sensory systems (Broom & Johnson 1993). Our current understanding of how intensity changes over time in response to acute and continued exposure to a hazard, or multiple hazards is improving, as more studies aim to quantify both the short- and long-term effects of different treatments on animal welfare (eg Willner *et al* 1992; Willner 2005; Jarvis *et al* 2006; Rutherford *et al* 2009; Jones *et al* 2010), although we could also turn to human medicine for models of pain intensity distributions over time (eg Menegazzi 1996).

Quantifying model parameters

Aven (2011) defines Quantitative Risk Assessment (QRA) as a process that:

> systemises the present state of knowledge including the uncertainties about the phenomena, processes, activities and systems being analysed. It identifies possible hazards/threats..., analyses their causes and consequences, and describes risk. A QRA provides a basis for characterising the likely impacts of the activity studied, for evaluating whether risk is tolerable or acceptable and for choosing the most effective and efficient risk policy, for example with respect to risk-reducing measures.

Although Aven (2011) was describing QRA, this definition would also apply equally to qualitative, or expert opinion (EO)-based risk assessment. The differences between the two approaches lie solely in the source and subsequent handling of the data (including estimating levels of uncertainty) — the overall aims and objectives remain the same for both.

Typically, data for welfare risk assessments are collected on each of the hazard and consequence characteristic factors, either directly from research published in peer-reviewed papers (eg Asher *et al* 2009; Summers *et al* 2010; Bettley *et al* 2012) or from experts in the field (EFSA 2008, 2009, 2010a,b). EO-based risk assessments have been the most common form of welfare risk assessment produced to-date. In part, this may be because most risk assessments concern a wide range of hazards with very large cascades of consequences, where systematic reviews would be extensive and time-consuming. It is therefore quicker and simpler to rely on the assimilated knowledge of a group of independent experts.

In addition, for many animal welfare problems, appropriate data do not exist, or exist only in part in the peer-reviewed literature to permit a comprehensive risk assessment. What effect might the type of data collected, be it qualitative or quantitative, have on the calculated estimate of risk? Although it is likely that EO-based and data-based estimates are similar for those factors that are relatively simple to quantify (eg duration of effects of consequences, prevalence, duration of exposure to hazard, probability of exposure to hazard), this has not been explicitly tested. In part, this may be because quantifying these factors is actually no trivial task and to do so on the large scale often demanded in risk assessments (across several regions or countries, each with individual prevalence estimates) would be potentially complex, requiring either the incorporation of stochasticity into the assessment, or creating a summary value that represents the distribution (but which also loses some of the variability inherent in the population). However, as mentioned previously, this is relatively simple compared with the vastly more difficult issue of quantifying intensity of consequences. This is, in effect, putting an exact value on the level of suffering caused. To-date, there is no single universal welfare indicator capable of reliably gauging welfare both across contexts and species. More often than not, different indicators point in different directions, making interpretation difficult (Rushen 1991; Mason & Mendl 1993; Mendl *et al* 2009; Nicol *et al* 2011). In terms of welfare risk assessment, this makes quantifying intensity using a single quantitative trait impossible at present. However, there are alternative means of assessing intensity that are quantitative. For example, in the Generic Illness Severity Index for Dogs (GISID) (Asher *et al* 2009), a suite of factors are scored to gauge the level of suffering experienced by the animal. In the GISID, each disorder is scored on four factors: prognosis, treatment, complications and behaviour. Each factor is given a score on a scale of zero to four, where zero is the least severe and four the most severe departure from a normal, healthy condition. The maximum possible GISID score of 16 is therefore given when all four factors have the maximum score of four. In both cases, this is based on a combination of owner and veterinary opinion of key behaviours affected, prognosis, complications and type of medical or surgical intervention possible. The prognosis score, developed originally to assess severity of conditions in horses (Bettley *et al* 2012), is a less detailed version of the GISID that is more widely applicable to other species. This version of the score focuses on just one factor, prognosis, as it was recognised that there was not enough information in the literature to score the other factors included in the original GISID. Both indices are applicable only to health-related welfare issues and are not suitable for use on the wide variety of hazards that cause mental suffering without physical signs.

A second question to consider in scoring intensity is how it changes over the course of a problem: does it increase,

https://doi.org/10.7120/096272812X13345905673764 Published online by Cambridge University Press

decrease, stay the same? For example, with some welfare hazards, such as high stocking density in commercial broiler houses, intensity would be predicted to increase over time as the birds grow larger and take up more space. Here, the consequences of the hazard would be relatively minor until the final few weeks of the broilers' lives when they approach slaughter weight. On the contrary, the intensity of other consequences may be expected to decrease over time. For example, a superficial injury such as a paper-cut on the finger is most painful shortly after it is inflicted, but the intensity decreases over time as the injury heals. The issue with changes in intensity and the fact that different individuals will experience different intensities for the same problem, means that again, as with prevalence, there is no single appropriate value that can be used to summarise across all individuals and all experiences. A more realistic method of incorporating intensity in risk assessments would be to include a probability distribution of intensity scores. This would require specific studies to be conducted to collect this type of data for animal welfare issues.

Quantifying and incorporating variability and uncertainty into risk assessments

For every value incorporated into a welfare risk assessment, be it intensity, duration, prevalence, probability of exposure to a hazard etc, there will be a level of uncertainty associated with it. As with any other area of scientific enquiry, we cannot be completely confident that the value we have measured is totally accurate with absolutely no flaws. Risk analysts argue that the assessment of risk is tantamount to the assessment of uncertainty. For example, Aven and Renn (2009) define risk as:

> uncertainty about and severity of the consequences (or outcomes) of an activity with respect to something that humans value.

However, in animal welfare risk assessments, to-date, there has been a general lack of second-order probability estimation (estimations of probabilities of probabilities — ie estimating uncertainty in a given value). Incorporating second-order probabilities into future risk assessment models could be achieved through a move from frequentist principles to a Bayesian framework (with the calculation of prior and updatable, posterior probabilities). The traditional risk assessment procedure, using frequentist principles, is typically more focused on assessing the chance of occurrence, and in this respect, uncertainty is incorporated as variation through the inclusion of a confidence interval for each assigned value.

Variation in the values given by experts in traditional EO-based risk assessments have been incorporated into some welfare risk assessments (Asher *et al* 2009; EFSA 2010a,b; Summers *et al* 2010), but on the whole this is not considered. Typically, each expert in a pool scores each hazard once, and different experts in the same pool will give different scores for the same factors. There is no investigation of either inter- or intra-expert scoring. It is possible that should the same group of experts be asked to complete the same survey six months into the future, a slightly different risk assessment would result. Similarly, a random sample of different experts may give different values for the same risk assessment. This is one of the key criticisms of EO-based risk assessment procedures (Vose 2008) and one of the key arguments for moving more to quantitative, data-based methods. However, although on the whole, the use of data is preferable, this does not render EO-based assessments futile. Rather, increasing the pool of experts so that it is as large as possible without compromising level of expertise and carefully including estimates of variation for each of the factors could lead to the development of powerful assessments.

Discussion

According to Slovic (1999), risk assessment is by its nature subjective, blending "science and judgement with psychological, social, cultural, and political factors". This may at first appear to be a rather damning statement with regards the use of risk assessment and its continued application to animal welfare problems, with the imperative requirement of objectivity. However, the extent to which it is true depends largely on the methods used to conduct the assessment. For animal welfare risk assessment, which as a field of research is still very much in its infancy, there have been a series of developments over recent years to make the process objective, and to make advances towards fully quantitative assessments based solely on data collected in scientific investigations. The major rate-determining step in this process has been the availability of accurate, unbiased, relevant data for the problems being assessed (eg Asher *et al* 2009; Collins *et al* 2010, 2011; Summers *et al* 2010). However, it is true to say that the risk assessment process has highlighted where the uncertainties are in our understanding of different animal systems and the welfare issues that arise from them. Aven (2011) argues that one of the principle benefits of quantitative risk assessment in general is that this process points very clearly at where the gaps in our knowledge lie. Scientific investigations require both reliability and validity. The tool must be consistent in its measurements (ie it must be reliable) and it must be measuring what it purports to be measuring (validity). In this paper, I have highlighted three key pitfalls of the risk assessment process as it is currently considered in relation to animal welfare problems. These issues are ones principally concerned with reliability, but they are not insurmountable. The development of more realistic risk assessment models may be the next step for animal welfare risk assessment. Although with increasing realism also comes greater potential for error as more assumptions are imposed and more parameters are measured with increased chance of measurement errors. So what is the solution? Firstly, to accept that risk assessment is a modelling process and is flawed in the same ways as any other model of real-world processes. However, just as with other models, we take from them what we can, but use them principally as a guide for further discussion and to highlight where further experimental work is needed.

Animal welfare implications

Risk assessment procedures have been used to support decision-making and prioritisation of welfare problems within and between species. Consideration of the potential shortcomings of this approach is therefore critical to more informed decision-making. Developing models that overcome these shortcomings will lead to more accurate assessment of the relative impacts of different problems on the welfare of animals.

Conclusion

Risk assessment methods are increasingly being used in animal welfare to rank different welfare issues relative to each other for prioritisation. Although these methods have a great deal of potential, the field is very much in its infancy and the limitations and pitfalls of current, simplistic models must be understood if we are to use their outcomes appropriately. This paper has discussed the three major issues with current welfare risk assessment methods that need to be considered and developed in future models: incorporation of the lack of independence between hazards and consequences, between the factors being scored and potentially between individual experts; quantifying the factors for hazard characterisation; and quantifying and incorporating the variability and uncertainty in scores into the models.

References

Anderson M, Jaykus L-A, Beaulieu S and Dennis S 2011 Pathogen-produce pair attribution risk ranking tool to prioritise fresh produce commodity and pathogen combinations for further evaluation (P3ARRT). *Food Control 22*: 1865-1872. http://dx.doi.org/10.1016/j.foodcont.2011.04.028

Asher L, Diesel G, Summers JF, McGreevy P and Collins LM 2009 Inherited defects in pedigree dogs I: disorders that are related to breed standards. *The Veterinary Journal 182*: 402-411. http://dx.doi.org/10.1016/j.tvjl.2009.08.033

Aven T 2011 *Quantitative Risk Assessment: The Scientific Platform*. Cambridge University Press: New York, USA

Aven T and Renn O 2009 On risk defined as an event where the outcome is uncertain. *Journal of Risk Research 12*: 1-11. http://dx.doi.org/10.1080/13669870802488883

Bettley CD, Asher L, Cardwell JM and Collins LM 2012 Inherited disorders in domestic horse breeds. *Animal Welfare 21*: 59-64

Broom DM and Johnson KG 1993 *Stress and Animal Welfare*. Kluwer Academic Publishers: Dordrecht, The Netherlands

Collins LM 2005 *Non-intrusive social preference indicators in broiler chickens*. DPhil Thesis, University of Oxford, Oxford, UK

Collins LM, Asher L, Summers JF and McGreevy PD 2011 Getting priorities straight: Risk assessment and decision-making in the improvement of inherited disorders in pedigree dogs. *The Veterinary Journal 189*: 147-154. http://dx.doi.org/10.1016/j.tvjl.2011.06.012

Collins LM, Asher L, Summers JF, Diesel G and McGreevy PD 2010 Welfare epidemiology as a tool to assess the welfare impact of inherited defects on the pedigree dog population. *Animal Welfare 19*: 67-75

EFSA (European Food Safety Authority) 2008 Animal welfare aspects of husbandry for farmed Atlantic salmon. Scientific Opinion of the panel on animal health and welfare. *EFSA Journal 736*: 1-31

EFSA (European Food Safety Authority) 2009 Scientific Opinion of the Panel on Animal Health and Welfare. Species-specific welfare aspects of the main systems of stunning and killing of farmed fish: rainbow trout. *EFSA Journal 1013*: 1-55

EFSA (European Food Safety Authority) 2010a Scientific Opinion on the influence of genetic parameters on the welfare and the resistance to stress of commercial broilers. *EFSA Journal 8*: 1666

EFSA (European Food Safety Authority) 2010b Scientific Opinion on welfare aspects of the management and housing of the grandparent and parent stocks raised and kept for breeding purposes. *EFSA Journal 8*: 1667

Jarvis S, Moinard C, Robson SK, Baxter E, Ormandy E, Douglas AJ, Seckl JR, Russell J and Lawrence AB 2006 Programming the offspring of the pig by prenatal social stress: neuroendocrine activity and behaviour. *Hormones and Behaviour 49*: 68-80. http://dx.doi.org/10.1016/j.yhbeh.2005.05.004

Jones MA, Mason G and Pillay N 2010 Early social experience influences the development of stereotypic behaviour in captive-born striped mice *Rhabdomys*. *Applied Animal Behaviour Science 123*: 70-75. http://dx.doi.org/10.1016/j.applanim.2009.12.009

Mason G and Mendl M 1993 Why is there no simple way of measuring animal welfare? *Animal Welfare 2*: 310-319

Mendl M, Burman OHP, Parker R and Paul E 2009 Cognitive bias as an indicator of animal emotion and welfare: emerging evidence and underlying mechanisms. *Applied Animal Behaviour Science 118*: 161-181. http://dx.doi.org/10.1016/j.applanim.2009.02.023

Menegazzi J 1996 Measuring pain at baseline and over time. *Annals of Emergency Medicine 27*: 433-435. http://dx.doi.org/10.1016/S0196-0644(96)70224-X

Nicol CJ, Caplen G, Edgar J, Richards G and Browne WJ 2011 Relationships between multiple welfare indicators measured in individual chickens across different time-periods and environments. *Animal Welfare 20*: 133-143

OED (Oxford English Dictionary) 2010 *Oxford English Dictionary, Third Edition*. Online Version June 2011. http://www.oed.com.ezproxy1.library.usyd.edu.au/view/Entry/166306. (Accessed 29 June 2011)

Palisade Corporation 2011 *@RISK Software*. Palisade Corporation: Middlesex, UK. www.palisade.com

Qu X, Meng Q and Li S 2011 Ship collision risk assessment for the Singapore Strait. *Accident Analysis and Prevention 43*: 2030-2036. http://dx.doi.org/10.1016/j.aap.2011.05.022

Rushen J 1991 Problems associated with the interpretation of physiological data in the assessment of animal welfare. *Applied Animal Behaviour Science 28*: 381-386. http://dx.doi.org/10.1016/0168-1591(91)90170-3

Rutherford KMD, Robson SK, Donald RD, Jarvis S, Sandercock DA, Scott EM, Nolan AM and Lawrence AB 2009 Pre-natal stress amplifies the immediate behavioural responses to acute pain in piglets. *Biology Letters 5*: 452-454. http://dx.doi.org/10.1098/rsbl.2009.0175

Slovic P 1999 Trust, emotion, sex, politics, and science: surveying the risk-assessment battlefield. *Risk Analysis 19*: 689-701. http://dx.doi.org/10.1111/j.1539-6924.1999.tb00439.x

https://doi.org/10.7120/096272812X13345905673764 Published online by Cambridge University Press

Summers JF, Diesel G, Asher L, McGreevy PD and Collins LM 2010 Inherited diseases in pedigree dogs II: Diseases not related to breed standards. *The Veterinary Journal 183*: 39-45. http://dx.doi.org/10.1016/j.tvjl.2009.11.002

Vose D 2008 *Risk Analysis: A Quantitative Guide, Third Edition.* John Wiley & Sons Ltd: Chichester, UK

Willner P 2005 Chronic Mild Stress (CMS) revisited: consistency and behavioural-neurobiological concordance in the effects of CMS. *Neuropsychobiology 52*: 90-110. http://dx.doi.org /10.1159/000087097

Willner P, Muscat R and Papp M 1992 Chronic mild stress-induced anhedonia: a realistic animal model of depression. *Neuroscience and Biobehavioral Reviews 16*: 525-534. http://dx.doi.org/10.1016/S0149-7634(05)80194-0

Zabeo A, Pizzol L, Agostini P, Critto A, Giove S and Marcomini A 2011 Regional risk assessment for contaminated sites Part 1: Vulnerability assessment by multicriteria decision analysis. *Environment International 37*: 1295-1306. http://dx.doi.org/10.1016/j.envint.2011.05.005