Implementing Welfare Quality® in UK assurance schemes: evaluating the challenges

CAE Heath*, Y Lin, S Mullan, WJ Browne and DCJ Main

School of Veterinary Sciences, University of Bristol, Langford House, Langford BS40 5DU, UK
* Contact for correspondence and requests for reprints: cheryl.heath@bristol.ac.uk

Abstract

This paper presents an account of a Welfare Quality® assessment of 92 dairy farms carried out by seven experienced assessors. The aim was to evaluate the potential of the Welfare Quality® assessment protocol with respect to its uptake by UK farm assurance schemes. Data collection, and measure aggregation were performed according to the Welfare Quality® protocol for dairy cows. This study examined the data itself, by the testing of how hypothetical interventions might be reflected in changes in the aggregated scores, and also investigated human-related aspects, through inter-assessor standardisation sessions to evaluate reliability, and an assessor focus group to collect feedback. Overall, three main ‘challenges’ were identified. The first challenge related to the large amount of missing data. Unexpectedly, this was such that it was only possible to calculate an overall classification for 7% of farms. The second challenge concerned the way in which aggregated scores did not always reflect hypothetical interventions. The final challenge was inter-assessor reliability, where not all assessors were found to achieve acceptable levels of agreement on a number of outcome measures by the third training session. Suggestions for managing these challenges included, follow-up to assessor training, the use of multiple imputation methods to fill in missing data, and, where applicable, not aggregating the scores. The conclusion of the study was that the protocol provided useful information from which to make an informed selection of measures, but that the challenges, combined with the lengthy assessment time, were too great for its use as a certification tool.

Keywords: animal welfare, dairy cow, focus group, on-farm assessment, score aggregation, Welfare Quality®

Introduction

Traditionally, animal welfare assessment schemes have been concerned with the measurement of inputs into the husbandry system, such as the provision of resources and aspects of the farm management system. These ‘input-based’ measurements have the advantage of being easy to measure and stable over time. Behavioural and physical observations of the animals can be understood to represent the outcome of the husbandry system and are referred to as outcome-based measurements. The Farm Animal Welfare Council (2005) has recommended the inclusion of outcome measures in farm assurance certification schemes. An advantage of an outcome-based assessment is that the animal is the focus of the assessment, and this allows comparisons across farming systems. It is for this reason that animal-based measurements are now considered to provide a more direct account of welfare, reflecting the experience of the animal (Webster et al 2004; Welfare Quality® 2009a). However, one of the criticisms of outcome-based measures is that they involve a degree of subjective interpretation, and the scoring is at risk of assessor bias. For this reason, standardisation of scores and adequate training are vital to minimise individual differences.

The question of what should be measured to assess welfare is often debated. While science can provide answers as to how things should be measured, the issue of what is considered important for animal welfare represents more of an ethical decision. Animal welfare is a multi-dimensional concept (Fraser 1995), encompassing both mental and physical health (Dawkins 2003) and, as such, it might be expected that a welfare assessment will reflect this in both the measures that are collected, and the manner in which they are aggregated. While an aim of welfare assessment may be to be as comprehensive as possible, outcome-based measures are inherently time consuming to collect, especially those relating to certain behavioural observations, given their infrequent displays. Whether the substantial amount of time required to collect the measures is warranted may depend on requirements and time restriction of the specific application of the protocol.

The Welfare Quality® assessment protocols present an extensive, scientifically robust, outcome-based account of welfare (Blokhuis et al 2010), whose conceptual underpinnings reflect the opinions of stakeholders from numerous backgrounds, including scientists, social scientists and the general public (Miele et al 2011). The protocols describe a