Repeated locomotion scoring of a sow herd to measure lameness: consistency over time, the effect of sow characteristics and inter-observer reliability

RB D’Eath

Animal Behaviour & Welfare, Veterinary Science Research Group, SAC, West Mains Road, Edinburgh EH9 3JG, UK; email: rick.death@sac.ac.uk

Abstract

Investigating variability of scores between different observers, between animals and over time aids the design of valid sampling methodologies for measuring animal welfare. Locomotion scores (0 to 5 scale) were collected: i) from 154 sows in one herd, using three to five observers each time, and scoring sows on up to ten occasions over a 19-month period; and ii) for 123 of these sows, locomotion scoring also took place prior to farrowing and at weaning. The distribution of scores was highly skewed towards low scores (0: 84.8%, 1: 9.5%, 2: 4.0%, 3+: 1.7%). Sows showed moderate consistency in their scores over time and later parity sows had higher scores, but there was no effect of stage in the reproductive cycle (days pregnant, pre-farrowing, post-weaning). This suggests that infrequent visits to a farm (eg annual) might provide an accurate estimate of the extent of lameness if a representative range of parities was sampled, although a larger study of more farms would be required to investigate this. The three different types of agreement between observers (absolute differences, matching and association) were assessed as follows: i) analysis of absolute differences between observers showed that the farm manager scored lower than researchers/technicians; ii) exact matching approaches suggested fair or good agreement — agreement was poorest for mild gait abnormalities (score 1 ‘stiff’), and agreement improved if scores were combined into ‘sound’ (0–1) and ‘lame’ (2–5) categories; and iii) measures of association suggested moderate agreement. Inter-observer reliability improved over time until the 5th scoring event. To improve inter-observer agreement, observer training/practice and the use of fewer categories are recommended, and inter-observer agreement should be checked regularly.

Keywords: animal welfare, inter-observer agreement, lameness, locomotion scoring, pigs, sows

Introduction

On-farm measurement of animal health and welfare is an important and current issue, to meet consumers’ demands for demonstrably high standards of farm animal welfare (Blokhuis et al 2003, 2008). Setting standards and inspecting to ensure that they are met is a goal of government agencies (Gibbens 2008) and of voluntary farm assurance schemes such as the UK Red Tractor scheme, free range, ecological and organic (Main et al 2003, 2007; Veissier et al 2008). Membership of such schemes depends generally on the producer meeting certain ‘design criteria’ (Rushen & de Passillé 1992) relating to the housing and resources provided to animals (such as stocking density, drinkers, substrates), management (eg weaning age in pigs, age at slaughter) and administration (eg keeping accurate records of the use of drugs). Conformance with these criteria is generally assessed in a visit which takes place approximately once a year and takes less than a day to complete (Main et al 2007).

With a few exceptions, direct assessment of health and welfare by inspecting the animals themselves (animal-based or ‘performance criteria’; Rushen & de Passillé 1992) has not formed part of these schemes. Recently, an EU-funded project ‘Welfare Quality®’ (Blokhuis et al 2003) developed a comprehensive animal-based scoring system for on-farm assessment of animal health and welfare for pigs (Welfare Quality® 2009) and other housed species. The measures adopted were assessed for validity (does the indicator really measure what it should), repeatability (across observers), and feasibility (can it be assessed quickly enough to be included in a short visit). This process has been described in general terms but not in detail (Keeling et al 2009; Knierim & Winckler 2009).

Integration of multiple measures into an overall assessment is a difficult part of Welfare Quality® (Botreau et al 2007a,b,c, 2009; Knierim & Winckler 2009) and of similar schemes (eg Main et al 2007). Even prior to reaching this stage, though, there are a number of difficulties (Knierim & Winckler 2009). For any single measure, there are already practical constraints: on-farm scoring of animal welfare involves a sampled subset of animals from each age class and housing type, often by one trained observer in an annual visit of less than one day (Mullan et al 2009). Some researchers have attempted to assess the effect of such low