J. Dairy Sci. 97:3523–3530
http://dx.doi.org/10.3168/jds.2013-7464

Evaluation of the sample needed to accurately estimate outcome-based measurements of dairy welfare on farm

M. I. Endres,*[1] K. M. Lobeck-Luchterhand,* L. A. Espejo,† and C. B. Tucker‡
*Department of Animal Science, and
†Department of Veterinary Population Medicine, University of Minnesota, St. Paul 55108
‡Department of Animal Science, University of California, Davis 95616

ABSTRACT

Dairy welfare assessment programs are becoming more common on US farms. Outcome-based measurements, such as locomotion, hock lesion, hygiene, and body condition scores (BCS), are included in these assessments. The objective of the current study was to investigate the proportion of cows in the pen or subsamples of pens on a farm needed to provide an accurate estimate of the previously mentioned measurements. In experiment 1, we evaluated cows in 52 high pens (50 farms) for lameness using a 1- to 5-scale locomotion scoring system (1 = normal and 5 = severely lame; 24.4 and 6% of animals were scored ≥3 or ≥4, respectively). Cows were also given a BCS using a 1- to 5-scale, where 1 = emaciated and 5 = obese; cows were rarely thin (BCS ≤2; 0.10% of cows) or fat (BCS ≥4; 0.11% of cows). Hygiene scores were assessed on a 1- to 5-scale with 1 = clean and 5 = severely dirty; 54.9% of cows had a hygiene score ≥3. Hock injuries were classified as 1 = no lesion, 2 = mild lesion, and 3 = severe lesion; 10.6% of cows had a score of 3. Subsets of data were created with 10 replicates of random sampling that represented 100, 90, 80, 70, 60, 50, 40, 30, 20, 15, 10, 5, and 3% of the cows measured/pen. In experiment 2, we scored the same outcome measures on all cows in lactating pens from 12 farms and evaluated using pen subsamples: high; high and fresh; high, fresh, and hospital; and high, low, and hospital. For both experiments, the association between the estimates derived from all subsamples and entire pen (experiment 1) or herd (experiment 2) prevalence was evaluated using linear regression. To be considered a good estimate, 3 criteria must be met: $R^2 > 0.9$, slope =1, and intercept = 0. In experiment 1, on average, recording 15% of the pen represented the percentage of clinically lame cows (score ≥3), whereas 30% needed to be measured to estimate severe lameness (score ≥4). Only 15% of the pen was needed to estimate the percentage of the herd with a hygiene score ≥3, whereas 30% to estimate the prevalence of severe hock lesions. Estimating very thin and fat cows required that 70 to 80% of the pen be measured. In experiment 2, none of the pen subsamples met our criteria for accurate estimates of herd prevalence. In conclusion, we found that both a higher percentage of the pen must be sampled to generate accurate values for relatively rare parameters and that the population measured plays an important role in prevalence estimates.
Key words: dairy welfare assessment, sampling size, lameness

INTRODUCTION

Animal welfare assessments and audits are becoming more common in the dairy industry. Examples of these assessments include the National Dairy FARM Program (www.nationaldairyfarm.com) in the United States and the Welfare Quality Program (www.welfarequality.net) in Europe. Outcome- or animal-based measurements, such as locomotion, hock lesion, hygiene, and BCS, are included in these assessments. Obtaining these animal-based measures is essential to truly assess the animal welfare quality on each farm, as these measurements provide valuable information about how cows are responding to their environment. Webster (2005) summarized that an effective animal welfare assessment program must incorporate both measures of husbandry provision (e.g., resources and management) and welfare (he defined welfare as capacity to sustain fitness and avoid suffering).

Ideally, sampling regimens used in each assessment provide accurate prevalence information for the previously mentioned outcome-based measurements that can evaluate welfare. The traditional approach of using sample size calculation to estimate the herd-level prevalence in a finite population depends on the population size, expected prevalence, level of precision, and the confidence level of the estimation (Cannon and Roe, 1982; Dohoo et al., 2010). Therefore, this approach assumes knowledge of the approximate prevalence before

Received September 6, 2013.
Accepted February 5, 2014.
[1]Corresponding author: miendres@umn.edu

sampling the cows, which may include high levels of uncertainty in field conditions. Relatively little is known about sampling required to ensure accuracy in animal welfare evaluations, as most previous studies have assessed the entire herd when evaluating (e.g., lameness prevalence; Huxley et al., 2004; Haskell et al., 2006). It is known, however, that scoring entire herds can be time and cost prohibitive, particularly on large farms.

Very limited research has been done on sampling methods for dairy welfare evaluation, especially in the United States, where 47% of herds have 500 or more lactating cows (USDA, 2008). Dairy welfare assessment programs calculate sample number of cows per pen based on the number of cows in the herd or pen (e.g., National Dairy FARM Program; Welfare Quality), whereas some industry professionals have suggested that perhaps a representative pen or subsamples of pens be used to estimate herd prevalence. Validation of these 2 approaches for all outcome-based measurements described has not been reported.

To date, 2 studies evaluating sampling methods for dairy cattle have focused on lameness. Main et al. (2010) investigated sampling strategies to monitor lameness in dairy cattle in 224 herds in the United Kingdom. They found that presence of more than 1 severely lame cow on a farm could be an indication of a herd lameness problem; for the 182 farms with at least 1 severely lame cow present, 80% of them had >25% lameness prevalence. They also suggested that using a sampling strategy based on milking parlor exit order could provide an accurate estimate of true prevalence in the herd. For 36 herds that were larger than 100 cows, sampling a maximum of 100 cows from the middle of the milking order produced an estimate of prevalence within 5% of the true prevalence (their criteria for a good estimate) on 83% of farms. In larger herds, however, this would require the presence of the evaluator on the farm for about 8 h to score every pen (fresh, high, mid, and low production strings) in the middle of milking order. Though valuable, the largest farm in Main et al. (2010) had 268 cows, some farms had only 1 group of cows, and the relationship between true and estimated prevalence was not statistically compared with farm as the experimental unit. A second study by Hoffman et al. (2013) used 5 herds in the western United States. Herd size ranged from 148 to 2,744 cows. They found that estimates of lameness prevalence using both cows in the middle of milking parlor exit and a calculated sample across the herd were within 5% of true values. However, with only 5 herds in total, they did not have adequate power to assess the accuracy of these methods across farms (where dairy operation serves as the experimental unit). They also investigated locomotion scoring certain pens or combinations of pens. This would allow the assessor to schedule the visit at the time those pens are being milked, reducing the overall welfare assessment time. They concluded that pen-level prevalence was variable and not a good predictor of herd-level lameness prevalence, again based on being within 5% of true values rather than a statistical comparison.

The objective of the present study was to statistically evaluate the proportion of cows in the high pen or subsamples of pens needed to provide an accurate estimate of locomotion, body condition, hygiene, and hock lesion scores.

MATERIALS AND METHODS

Data Sets

Two previously collected, on-farm data sets were used in the current study. In experiment 1, data from 52 high group pens in 50 farms were used. Dairy farms for this data set had been selected randomly from the total population of herds having more than 150 cows situated within a geographical area where most of the dairy farms in Minnesota are located. All farms had Holstein cows as the predominant breed. No previous knowledge existed about prevalence of lameness, hock lesions, hygiene, or BCS in these farms. Farms had, on average, 474 ± 321 (mean ± SD) cows and milk production averaged 32.9 ± 5.4 kg of FCM per cow per day. Group size averaged 117 ± 51 cows. A total of 5,626 cows were included in the data set. Sixteen of the 52 total groups in the data set were high-production groups, which did not include first-lactation cows. High-production cows (as defined by the herd manager) housed in the hospital pen during the visit also were included in the data set. Approximately 49% of the cows included in the data set were more than 150 DIM. Thirty percent of the pens had deep-sand freestalls and 70% had mattresses.

In experiment 2, data from 12 freestall dairy farms in Minnesota and southeast South Dakota were used. Dairy farms for this data set had Holstein as the primary breed and 2 herds had approximately 30% Jersey-Holstein crossbreds. All barns used deep-bedded sand stalls, and freestalls were similar in size and design across facilities, with an average width of approximately 122 cm, length of 244 cm, and neck rail height of 122 cm. Farms were visited once each season for a total of 4 visits per farm. Visits during January and February were considered winter; April and May, spring; July and August, summer; and October and November, fall. The mean number of lactating animals on each farm ranged from 399 to 1,564, with an average of 913 cows per farm. Average production was 37.3 ± 12.0 kg of FCM/cow per day. A total of 42,693 cows were scored across all farms and visits.