J. Dairy Sci. 98:1–12
http://dx.doi.org/10.3168/jds.2014-8255

Optimization of reproductive management programs using lift chart analysis and cost-sensitive evaluation of classification errors

Saleh Shahinfar,*[1] Jerry N. Guenther,* C. David Page,† Afshin S. Kalantari,* Victor E. Cabrera,* Paul M. Fricke,* and Kent A. Weigel*
*Department of Dairy Science, and
†Department of Biostatistics and Medical Informatics and Department of Computer Science, University of Wisconsin, Madison 53706

ABSTRACT

The common practice on most commercial dairy farms is to inseminate all cows that are eligible for breeding, while ignoring (or absorbing) the costs associated with semen and labor directed toward low-fertility cows that are unlikely to conceive. Modern analytical methods, such as machine learning algorithms, can be applied to cow-specific explanatory variables for the purpose of computing probabilities of success or failure associated with upcoming insemination events. Lift chart analysis can identify subsets of high fertility cows that are likely to conceive and are therefore appropriate targets for insemination (e.g., with conventional artificial insemination semen or expensive sex-enhanced semen), as well as subsets of low-fertility cows that are unlikely to conceive and should therefore be passed over at that point in time. Although such a strategy might be economically viable, the management, environmental, and financial conditions on one farm might differ widely from conditions on the next, and hence the reproductive management recommendations derived from such a tool may be suboptimal for specific farms. When coupled with cost-sensitive evaluation of misclassified and correctly classified insemination events, the strategy can be a potentially powerful tool for optimizing the reproductive management of individual farms. In the present study, lift chart analysis and cost-sensitive evaluation were applied to a data set consisting of 54,806 insemination events of primiparous Holstein cows on 26 Wisconsin farms, as well as a data set with 17,197 insemination events of primiparous Holstein cows on 3 Wisconsin farms, where the latter had more detailed information regarding health events of individual cows. In the first data set, the gains in profit associated with limiting inseminations to subsets of 79 to 97% of the most fertile eligible cows ranged from $0.44 to $2.18 per eligible cow in a monthly breeding period, depending on days in milk at breeding and milk yield relative to contemporaries. In the second data set, the strategy of inseminating only a subset consisting of 59% of the most fertile cows conferred a gain in profit of $5.21 per eligible cow in a monthly breeding period. These results suggest that, when used with a powerful classification algorithm, lift chart analysis and cost-sensitive evaluation of correctly classified and misclassified insemination events can enhance the performance and profitability of reproductive management programs on commercial dairy farms.

Key words: machine learning, reproductive management, lift chart analysis, cost-sensitive evaluation, dairy cattle

INTRODUCTION

The underlying physiological factors that determine the success or failure of an insemination event in dairy cows are complex and often unknown. Conception rate has a major effect on farm profitability because it affects the number of replacement heifers needed to maintain herd size, the capacity to generate and sell extra heifers, the proportion of pregnant cows in the herd, average milk production of these cows, insemination and veterinary costs, involuntary culling rate, and other factors (Britt, 1985; Meadows et al., 2005; Giordano et al., 2011).

Previous studies have attempted to predict the outcome of an insemination event in lactating dairy cows based on results of ovarian palpation (Ludwick and Rader, 1967), interactions between stage of lactation and pregnancy status (Sharma et al., 1990), or levels of NEFA and blood glucose (Garverick et al., 2013). Other studies have considered reproductive management programs from an economic point of view; for example, by estimating the economic value of a pregnancy (De Vries, 2006), finding an economically optimal voluntary waiting period (Inchaisri et al., 2011), or carrying out

Received April 19, 2014.
Accepted February 23, 2015.
[1]Corresponding author: shahinfar@wisc.edu

an economic comparison of natural service and timed AI programs (Lima et al., 2010; Giordano et al., 2012).

On most commercial dairy farms, it is common practice to inseminate all cows that are eligible for breeding, with the hope that modest conception rates coupled with high service rates will lead to efficient overall reproductive performance. This strategy may be economically sensible when insemination costs, specifically semen and technician service, are low. However, evidence that this strategy maximizes farm profitability is lacking, and insemination costs can be relatively high if sex-enhanced semen is used. Several studies have indicated that targeted reproductive management of specific (groups of) cows could be more profitable (e.g., Giordano et al., 2012), but no study to date has attempted to predict the outcomes of insemination events for individual cows before the breeding while subsequently evaluating the benefits and costs of correct and incorrect breeding decisions, respectively.

Shahinfar et al. (2014b) evaluated the classification accuracy of several machine learning algorithms when predicting the outcome of an insemination event for an individual cow based on the production, reproduction, health, and genetic information available before the insemination. The proportions of insemination events that were classified correctly as successes or failures were 72.3% for primiparous cows and 73.6% for multiparous cows in 5-fold cross-validation using a random forest algorithm. Classification accuracies of this magnitude bring into question the practice of inseminating every eligible cow, while ignoring the vast amount of information available about each cow's genetic potential, health history, and lactation performance. Furthermore, we can carry out a lift chart analysis (Witten and Frank, 2005) to identify a subset of cows with much greater probability of conception than the entire pool of eligible cows. Although such an analysis would, for example, facilitate the use of expensive sex-enhanced or high-genetic-merit semen for mating the subset of cows most likely to conceive, it would be limited by the fact that evaluation of algorithms based only on classification accuracy implies that the benefits and costs associated with all correctly classified or misclassified events are equal. In reality, the costs and benefits may differ considerably. For example, if insemination costs are low, the cost of failing to inseminate a cow that would have conceived (i.e., a false negative, **FN**) is likely to be greater than the cost of inseminating a cow that will not conceive (i.e., a false positive, **FP**). Moreover, the benefit derived from inseminating a cow that will conceive (i.e., a true positive, **TP**) may be greater than the benefit associated with failing to inseminate a cow that would not have conceived (i.e., a true negative, **TN**).

The objective of this study was to build upon the work of Shahinfar et al. (2014b), which showed that relatively high accuracy of classifying insemination events as successes or failures could be achieved before insemination. Lift chart analysis was used to identify subsets of cows with high probability of conception, and then cost-sensitive evaluation was used to assess the economic benefits that might be achieved. Implementing such an approach for reproductive management of lactating dairy cows on commercial farms can lead to the development of an on-farm decision support tool that can be used by reproductive management consultants or incorporated into herd management software programs.

MATERIALS AND METHODS

Data

Two data sets were used in the analyses presented herein. The first, which will be denoted as Data_1, consisted of 54,806 insemination events from 22,210 primiparous Holstein cows that were inseminated between 2000 and 2010 in 26 Alta Genetics (Watertown, WI) Advantage Program herds located throughout Wisconsin. These breeding records represented a subset of the data used by Shahinfar et al. (2014b), in which insemination events corresponded to 3 intervals of DIM: 60–90, 90–120, or 120–150, and 3 ranges of relative milk yield (**RMY**): between 18 and 6% below the within-herd mean, between 6% below and 6% above the within-herd mean, or between 6 and 18% above the within-herd mean. Mean conception rate in Data_1 was 32%, and individual herds ranged from 18 to 44%. Additional details about genetic predisposition, health history, and lactation performance of herds in Data_1 are provided in Shahinfar et al. (2014b).

The second, which will be denoted as Data_2, consisted of 17,197 insemination events from 5,356 primiparous Holstein cows that were inseminated between 2002 and 2013 in 3 additional herds also located in Wisconsin. These herds were chosen because they had more complete data regarding common early postpartum health disorders (i.e., mastitis, ketosis, lameness, displaced abomasum, metritis, and retained placenta) than the herds in Data_1. In contrast to Data_1, because of the small size of the data set and the need for a more general model, insemination events in Data_2 were not stratified by DIM or RMY, and all insemination events were analyzed jointly. The range in DIM at time of insemination was 12 to 479 in Data_2. Mean across-herd conception rate in Data_2 was 42%, and means for individual herds ranged from 38 to 45%.