



Benchmarking dairy herd health status using routinely recorded herd summary data

K. L. Parker Gaddis,^{*1} J. B. Cole,[†] J. S. Clay,[‡] and C. Maltecca[§]

^{*}Department of Animal Sciences, University of Florida, Gainesville 32611

[†]Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705-2350

[‡]Dairy Records Management Systems, Raleigh, NC 27603

[§]Department of Animal Science, North Carolina State University, Raleigh 27695

ABSTRACT

Genetic improvement of dairy cattle health through the use of producer-recorded data has been determined to be feasible. Low estimated heritabilities indicate that genetic progress will be slow. Variation observed in lowly heritable traits can largely be attributed to nongenetic factors, such as the environment. More rapid improvement of dairy cattle health may be attainable if herd health programs incorporate environmental and managerial aspects. More than 1,100 herd characteristics are regularly recorded on farm test-days. We combined these data with producer-recorded health event data, and parametric and nonparametric models were used to benchmark herd and cow health status. Health events were grouped into 3 categories for analyses: mastitis, reproductive, and metabolic. Both herd incidence and individual incidence were used as dependent variables. Models implemented included stepwise logistic regression, support vector machines, and random forests. At both the herd and individual levels, random forest models attained the highest accuracy for predicting health status in all health event categories when evaluated with 10-fold cross-validation. Accuracy (SD) ranged from 0.61 (0.04) to 0.63 (0.04) when using random forest models at the herd level. Accuracy of prediction (SD) at the individual cow level ranged from 0.87 (0.06) to 0.93 (0.001) with random forest models. Highly significant variables and key words from logistic regression and random forest models were also investigated. All models identified several of the same key factors for each health event category, including movement out of the herd, size of the herd, and weather-related variables. We concluded that benchmarking health status using routinely collected herd data is feasible. Nonparametric models were better suited to handle

this complex data with numerous variables. These data mining techniques were able to perform prediction of health status and could add evidence to personal experience in herd management.

Key words: herd health status, producer-recorded data, prediction, benchmarking

INTRODUCTION

To fully understand complex diseases, it is important to understand relationships between genotype, environment, and phenotype. Complex causal relationships have been identified between different diseases, culling, and production (Dhakal et al., 2015). Increased production of dairy cattle has resulted in a subsequent decline in health and fertility traits (Esposito et al., 2014). Concurrently, concern over animal welfare and use of antibiotics has steadily increased (Nyman et al., 2007). Understanding these relationships may help us to better describe the disease process (Dhakal et al., 2015). Genetic improvement of dairy cattle health has been determined to be feasible using producer-recorded data by several studies (Zwald et al., 2004; Parker Gaddis et al., 2014, 2012). Low estimated heritabilities of health events indicate, however, that genetic progress will be slow.

Variance observed in lowly heritable traits can largely be attributed to nongenetic or environmental factors. In typical genetic evaluations, adjustments for environmental effects are accomplished by considering them fixed effects. This disregards effects of management and environmental conditions on genetic expression (Windig et al., 2005). It also ignores any associations that exist between genetic and environmental effects. Dechow and Goodling (2008) showed that heritabilities estimated using data from high-performing herds were higher than those from typical or poor-performing herds, suggesting that rates of genetic gain may be higher when cows are provided with favorable environments for production. In addition, research has indicated that genetic correlations, such as between fertility and milk

Received May 20, 2015.

Accepted September 25, 2015.

¹Corresponding author: klpgaddis@ufl.edu

production, will depend upon herd environment (Windig et al., 2006). The question then arises as to whether more rapid improvement can be achieved if herd health programs incorporate environmental and managerial aspects.

Previous studies have investigated the effect of environmental characteristics on dairy cattle health. An early study was able to establish 5 farm “health profiles” according to the incidence levels of health disorders and farm structure data (Faye, 1992). Health disorders included infectious diseases of the foot, uterus, and teat, and calving disorders; farm structure was represented as traditional, intensive, or intermediate. Data were collected throughout 1979 from 83 dairy farms in France and included 25 specific health events in addition to herd management variables. Hierarchical classification was used to group the farms into similar classes and confirmed a relationship between farm type and herd health profile (Faye, 1992). Path analysis and multiple logistic regression were utilized to evaluate interrelationships between herd management practices and postpartum health disorders on 32 farms located in New York State (Correa et al., 1990). Disorders included dystocia, retained placenta, metritis, cystic ovary, milk fever, ketosis, left displaced abomasum, and mastitis. Management characteristics were collected through a questionnaire provided to the person primarily responsible for care of the herd. A 2-stage analysis was performed to identify management factors and develop a path model of interrelationships between herd management and herd incidence rate (Correa et al., 1990).

More recent studies have been conducted incorporating herd characteristics in relationship to reproductive efficiency (Löf et al., 2007; Schefers et al., 2010), production (Windig et al., 2005, 2006; Simensen et al., 2010), and health (Svensson et al., 2006; Green et al., 2007; Stengärde et al., 2012). Many of these studies have utilized surveys or questionnaires to assess herd characteristics (Correa et al., 1990; Sato et al., 2008; Hill et al., 2009), which can limit the amount of data that can be collected. Data collected from a designed study may not always reflect common management practices, thus limiting applicability (Coppa et al., 2013). Data can also be limited by the chosen analysis method. The majority of previous studies have used parametric statistical models to analyze herd characteristics (Svensson et al., 2006; Löf et al., 2007; Stengärde et al., 2012), which can suffer from problems with multiple testing and collinearities with numerous variables (Sato et al., 2008). Alternatively, nonparametric methodologies have recently been investigated, such as principal component analysis (Windig et al., 2006) or common

factor analysis (Enevoldsen et al., 1996), as well as regression-based decision trees (Schefers et al., 2010) to better handle numerous variables.

Farm staff or DHIA technicians report numerous herd characteristics regularly on farm test days. These reports include data on herd production, reproduction, genetics, udder health, and feed costs (Dairy Records Management Systems, 2014). Additional environmental data can be accessed through online databases such as the National Climatic Data Center (www.ncdc.noaa.gov), the United States Census Bureau (www.census.gov), and the United States Geographical Survey (www.usgs.gov). The availability of numerous variables from field data presents analysis challenges ranging from increased data preprocessing to increased computing time. Although the majority of prior research has been conducted with parametric statistical methods (Windig et al., 2005, 2006; Sato et al., 2008), a more flexible approach might be possible when analyzing large numbers of variables utilizing data mining techniques. Data mining allows patterns to be explored and is increasingly employed because of the explosion of data availability in many fields (Sullivan, 2012). The objective of this study was to utilize parametric and nonparametric methods to explore prediction of herd health status. Routinely collected herd summary data were used for benchmarking health status at the individual and herd level.

MATERIALS AND METHODS

Data

The DHI-202 Herd Summary provides a “comprehensive herd analysis and management report including production, reproduction, genetics, udder health, and feed cost information” (Dairy Records Management Systems, 2014). Data are collected by farm staff or DHI technicians and compiled each test day. Categories of data include production, income, and feed cost summary; miscellaneous herd information; reproductive summary of current breeding herd; reproductive summary of total herd; birth summary; yearly reproductive summary; cows to be milking, dry, or calving by month; stage of lactation profile; identification and genetic summary; production by lactation summary; current SCC summary; dry cow profile; yearly summary of cows entered and left the herd; and yearly production and mastitis summary. An example DHI-202 Herd Summary report is included in the supplementary material (Supplementary Figure S1; <http://dx.doi.org/10.3168/jds.2015-9840>); detailed information on the report can be found at www.drms.org. Data were

Download English Version:

<https://daneshyari.com/en/article/10973562>

Download Persian Version:

<https://daneshyari.com/article/10973562>

[Daneshyari.com](https://daneshyari.com)