



J. Dairy Sci. 101:1–11
<https://doi.org/10.3168/jds.2017-12999>
 © American Dairy Science Association®, 2018.

The effect of using cow genomic information on accuracy and bias of genomic breeding values in a simulated Holstein dairy cattle population

E. Dehnavi,*† S. Ansari Mahyari,*¹ F. S. Schenkel,† and M. Sargolzaei†‡

*Department of Animal Science, College of Agriculture, Isfahan University of Technology, Isfahan, 84156-83111, Iran

†Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, ON N1G 2W1, Canada

‡Semex Alliance, Guelph, ON N1H 6J2, Canada

ABSTRACT

Using cow data in the training population is attractive as a way to mitigate bias due to highly selected training bulls and to implement genomic selection for countries with no or limited proven bull data. However, one potential issue with cow data is a bias due to the preferential treatment. The objectives of this study were to (1) investigate the effect of including cow genotype and phenotype data into the training population on accuracy and bias of genomic predictions and (2) assess the effect of preferential treatment for different proportions of elite cows. First, a 4-pathway Holstein dairy cattle population was simulated for 2 traits with low (0.05) and moderate (0.3) heritability. Then different numbers of cows (0, 2,500, 5,000, 10,000, 15,000, or 20,000) were randomly selected and added to the training group composed of different numbers of top bulls (0, 2,500, 5,000, 10,000, or 15,000). Reliability levels of de-regressed estimated breeding values for training cows and bulls were 30 and 75% for traits with low heritability and were 60 and 90% for traits with moderate heritability, respectively. Preferential treatment was simulated by introducing upward bias equal to 35% of phenotypic variance to 5, 10, and 20% of elite bull dams in each scenario. Two different validation data sets were considered: (1) all animals in the last generation of both elite and commercial tiers ($n = 42,000$) and (2) only animals in the last generation of the elite tier ($n = 12,000$). Adding cow data into the training population led to an increase in accuracy (r) and decrease in bias of genomic predictions in all considered scenarios without preferential treatment. The gain in r was higher for the low heritable trait (from 0.004 to 0.166 r points) compared with the moderate heritable trait (from 0.004 to 0.116 r points). The gain in accuracy in scenarios with a lower number of training bulls was relatively

higher (from 0.093 to 0.166 r points) than with a higher number of training bulls (from 0.004 to 0.09 r points). In this study, as expected, the bull-only reference population resulted in higher accuracy compared with the cow-only reference population of the same size. However, the cow reference population might be an option for countries with a small-scale progeny testing scheme or for minor breeds in large counties, and for traits measured only on a small fraction of the population. The inclusion of preferential treatment to 5 to 20% of the elite cows led to an adverse effect on both accuracy and bias of predictions. When preferential treatment was present, random selection of cows did not reduce the effect of preferential treatment.

Key words: prediction accuracy, cow genomic data, preferential treatment, regression coefficient

INTRODUCTION

Genomic selection has become a common practice in dairy cattle breeding programs due to an ever-decreasing cost of high-throughput genotyping products (Scheffers and Weigel, 2012). Selection decisions in dairy cattle breeding programs are now being made based on genomic estimated breeding values (**GEBV**), which usually are available early in life and are more accurate than traditional parent average (Goddard and Hayes, 2009). This has resulted in shortening the generation interval significantly and consequently increasing the annual genetic gain.

In some countries, genomic evaluation is mainly based on a bull training population because of the higher accuracy of bull proofs. One problem with a bull-only training population is selection bias due to intense selection on bulls (Patry and Ducrocq, 2011; Schaeffer, 2014). In addition, the size of the training population and accuracy of observations are the most important factors determining the accuracy of genomic predictions (Goddard, 2009). Other factors, such as the extent of linkage disequilibrium (**LD**) and genetic architecture of the trait, play an important role as well (Goddard and Hayes, 2009). For GEBV of young

Received April 8, 2017.

Accepted December 17, 2017.

¹Corresponding author: s.ansari@cc.iut.ac.ir

animals, Wiggans et al. (2010) reported average reliability of 76% for production traits based on a training population of 7,173 Holstein proven bulls. However, for a smaller training population, which consisted of 3,576 proven bulls, an average expected reliability of 69% and observed reliability of 62 to 66% for production traits were reported by VanRaden et al. (2009). Therefore, countries with small bull training populations may have to cope with a lower accuracy of genomic predictions. Sharing genotypes and phenotypes with other countries to increase the size of training population seems to be a useful solution (VanRaden and Sullivan, 2010; Weller et al., 2015, 2017). However, collaboration with other countries increases the risk of genotype-environment interaction. The genotype-environment interaction is mainly due to genes that are differently expressed in different environments, leading to different SNP effects in each environment. An alternative way to increase the size of a training population is to include cow information (genotype and phenotype; Wensch-Dorendorf et al., 2011; Pryce et al., 2012). However, cow information needs to be incorporated cautiously because there might be bias due to preferential treatment (**PT**) of elite cows. Instead, using a randomly selected group of females or adjusting for potential PT bias could be beneficial. Moreover, there is a lower genetic variation among the bulls, because of more intense selection on sires. Therefore, adding cow information may provide some additional variation that can increase accuracy and reduce bias and, hence, may allow for potential young candidates to be ranked better for selection.

Generally, adding cow information to the training population is challenging, as results appear to vary depending on the implementation strategy used (Gao et al., 2015). Moreover, few studies have examined the effect of PT on the accuracy and bias of genomic breeding values. The benefit of using cow information in the training population mainly depends on the degree of PT, the size of the cow training population, the size of the bull training population, and relationships with candidate females and the heritability of the phenotype. Therefore, the objectives of present study were to (1) investigate the effect of adding cow information into a training population on the accuracy and bias of genomic predictions and (2) assess the effect of PT bias when cow information is added to the training population, using a simulated Holstein cattle population.

MATERIALS AND METHODS

Genome Structure

A genome consisting of 29 autosomes each with different lengths similar to the bovine chromosomes for

a total length of 2,496 cM was simulated (Bohmanova et al., 2010). Genotypes were simulated for 50,000 bi-allelic markers and 750 multi-allelic QTL (2, 3, or 4 alleles) with initial equal allele frequencies, and random location on the genome. The effects of QTL alleles were sampled from a gamma distribution with a shape parameter of 0.4. The mutation rate for both SNP and QTL was set at 1×10^{-4} under the recurrent mutation model (Sargolzaei and Schenkel, 2009). Interference rate was set to 25% (Weeks et al., 1994; Tapper et al., 2002; Sargolzaei and Schenkel, 2009). The ascertainment bias in the SNP panel was introduced by selecting segregating loci according to MAF distribution observed in North American bull data.

Population Structure

A Holstein dairy cattle population was simulated using QMSim software (Sargolzaei and Schenkel, 2009). Supplemental Figure S1 (<https://doi.org/10.3168/jds.2017-12999>) presents the simulation schematic. First, a historical population was generated to reach mutation-drift equilibrium. To mimic AI and progeny test breeding schemes and to establish long-range LD, an AI population was simulated. This population mimicked a 4-pathway dairy cattle selection program with the use of AI technology. After establishing the proper level of LD, a most recent population, consisting of the elite (i.e., nucleus) and commercial groups, was simulated for 6 generations under the genomic selection scheme (**GS** population).

In the GS population, the selection of young animals was based on a predefined reliability of 70%, which represents the squared accuracy of genomic selection in the United States (VanRaden et al., 2009; Wiggans et al., 2010). Because a proven bull EBV is based on a large number of daughters, it is more accurate than a cow EBV based on a limited number of daughters and her own record. Therefore, pseudo-phenotypes for bulls and dams were simulated by calculating different predefined accuracies based on de-regressed EBV in Holstein cattle in Canada. Data were provided by Canadian Dairy Network and consisted of 7,500 domestically proven bulls and 5,000 cows. The bulls were those used in the training set of the national Canadian genomic evaluation. To simulate pseudo-phenotypes with a predefined accuracy, a correlated vector to true breeding values (**TBV**) was simulated as follows:

$$\mathbf{V} = \begin{pmatrix} var & cov \\ cov & var \end{pmatrix}$$

$$cholesky(\mathbf{V}) = \mathbf{CC}'$$

Download English Version:

<https://daneshyari.com/en/article/8501133>

Download Persian Version:

<https://daneshyari.com/article/8501133>

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