Impact of preferential treatment of elite cows on accuracy of genomic predictions in a simulated dairy cattle population

E. Dehnavi1,2, S. Ansari Mahyari1, F. S. Schenkel2 & M. Sargolzaei2,3

1 Isfahan University of Technology, College of Agriculture, Department of Animal Science, P.O. Box 84156-83111, Isfahan, Iran
e.dehnavi@ag.tut.ac.ir mailto:Riberc@univit.com
mailto:edehnavi@uoguelph.ca(Corresponding Author)

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

3 Semex Alliance, Guelph, ON NIH 6J2, Canada
msargol@uoguelph.ca mailto:edehnavi@uoguelph.ca(Corresponding Author)

Summary

Implementation of genomic selection using a cow training population is of interest for countries with no or limited progeny-tested bull data. However, preferential treatment of elite cows may cause bias in genomic predictions. The objective of this study was to investigate the impact of using preferentially treated cows in the training population on accuracy and bias of genomic predictions. A population undergoing a four-pathway selection strategy similar to that in dairy cattle was simulated. Two traits with low (0.05) and moderate (0.3) heritability were considered. The training population consisted of between 2,500 and 20,000 randomly selected cows. Preferential treatment (PT) was simulated and introduced to 5, 10 and 20% of elite cows. Preferential treatment of elite cows resulted in lower accuracy of predictions compared to the control scenario without PT. The accuracy of genomic predictions in the control scenario ranged from 0.72 to 0.83 and from 0.75 to 0.86 for traits with heritability of 0.05 and 0.3, respectively. When the training population included 20% preferentially treated cows, corresponding accuracies decreased to 0.68 to 0.80 and 0.64 to 0.77, respectively. Training on cows with PT resulted in upward bias of genomic predictions (regression coefficient 0.79 and 0.45 for traits with heritability of 0.05 and 0.3, respectively). Generally, using cow data in the training population is an attractive way to implement genomic selection for countries with no or limited progeny-tested bull data. However, further investigation is needed to adjust for or remove bias due to potentially presence of preferential treatment.

Keywords: accuracy, bias, genomic selection, preferential treatment

Introduction

Genomic selection has been integrated in breeding programs, especially in dairy cattle (VanRadan et al., 2009). However, some countries lack a progeny test program and therefore they do not have access to highly reliable proofs on bulls to implement genomic selection. One possibility in these countries is to genotype a large number of phenotyped cows for the training population. One potential issue with cow data is bias due to preferential treatment (Wiggans et al., 2011). Preferential treatment defined as any management practice that increases production and applies on some of cows but not to their herd mates. The management practice can be better feeding, separate housing or greater number of days open (Khun et al., 1994). Primary existence of preferential treatment bias detected when there was
inconsistency between the parent average breeding value of a bull and performance of the bull's first crop daughters (Van Vleck, 1987). Some studies support the idea that preferential treatment occurs among elite cows, (Dassonneville et al. 2012; Murphy et al., 1982) while some did not report any evidence of preferential treatment (Graham et al., 1991). Therefore, the objective of this study was to assess the impact of preferential treatment of elite cows on accuracy and bias of genomic predictions.

Materials and methods

A genome consisting of 29 autosomes for a total length of 2,496 cM was simulated. 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 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).

A Holstein dairy cattle population was simulated using the QMSim software (Sargolzaei and Schenkel, 2009). First, a historical population followed by a recent AI breeding population was simulated to establish a level of linkage disequilibrium similar to that observed in dairy cattle. In the historical population, 1600 unrelated individuals were mated randomly for 4000 generation under random union of gametes followed by 10 more generations to expand the size of population to 40100 individuals (100 males and 40,000 females). The recent AI population consisted of elite and commercial groups which represented nucleus and commercial herds with considering some young bulls for progeny testing program. Selection of young animals and proven bulls were based on the predefined reliabilities of 40 and 90% which represented the reliabilities of parent average and proven bulls, respectively. Then, a main breeding population was created over 6 generations. The main population was under four-pathway selection that exploited genomic section. Selection of young animals was based on predefined reliability of 70% representing reliability of genomic selection especially for the US (VanRaden et al., 2009). Two traits with low (0.05) and moderate (0.3) heritability were considered. The phenotypic variance was set to 100. The simulation was replicated 20 times.

Preferential treatment was attributed to management effects (Khun et al., 1994; Weigel et al., 1994). Preferential treatment was simulated by summing to the phenotypes of 5, 10 or 20% of elite cows random numbers from a normal distribution with mean equal to 0.35*Vphen (≈ 28.5 and 4.8 for heritability of 0.3 and 0.05, respectively) and variance equal to (0.35*Vphen/3)². The training population included 2,500 to 20,000 cows as well.

Generation 5 was used as the training group and all animals born in generation 6 were used as the validation group. Genomic predictions were computed using the snpBLUP method implemented in the SNP1101 software (Sargolzaei, 2014). The genomic evaluation was carried out using the following model:

\[ y_i = \mu + W_{ij}u_i + \sum_{j=1}^{p} \beta_j X_{ij} + e_i \]

where \( y_i \) is the observation for individual \( i \), \( \mu \) is the overall mean, \( u_i \) is the random polygenic effect of individual \( i \), \( \beta_j \) is the regression coefficient for SNP \( j \), \( e_i \) is the residual, \( W_{ij} \) is an incidence matrix that links records to individual \( i \), and \( X_{ij} \) is \( (Z_{ij} - 2f_j)/\sqrt{2\sum_{j=1}^{p}(f_j(1-f_j))} \) where \( f_j \) is frequency of allele 1 for SNP \( j \) and \( Z \) is a
genotype matrix for \( n \) animals and \( p \) SNPs. Accuracy was assessed by Pearson correlation of estimated direct genomic values (DGVs) with true breeding values (TBVs). Bias was assessed by the slope of the linear regression of the TBVs on DGVs.

**Results and Discussion**

Figure 1 depicts the effects of PT and the size of cow training population on prediction accuracy. When there was no PT, accuracy ranged for different numbers of cows from 0.72 to 0.83 when heritability was 0.05 and from 0.75 to 0.86 when heritability was 0.30. In the presence of PT for 20\% of elite cows, accuracy decreased to 0.68-0.80 and to 0.64-0.77 for heritability of 0.05 and 0.30, respectively. Without PT, the regression coefficient of TBVs on DGVs was on average 1.14 and 1.04 when heritability was 0.05 and 0.30, respectively. The average regression coefficient of TBVs on DGVs declined to 0.79 and 0.45 with heritability of 0.05 and 0.30, respectively, when PT was introduced to 20\% of elite cows. The effect of preferential treatment was more when heritability was 0.30 than 0.05. This may related to construction of preferential treatment, i.e. generating random numbers according to 35\% of phenotypic variance. Although, phenotypic variance for two traits was similar in the first generation of simulated population, it changed over generations by selection. Increasing the number of individuals in the training population resulted in an increase of non-biased information and hence a reduction of the effect of preferential treatment.

When it is aimed to construct the training population using cows, bull dams are the first target for exploitation. The use of potentially biased records of these genotyped elite cows in genomic evaluation may have two major effects one on the DGV of these cows and their relatives and another on the prediction equations. Dassonneville et al. (2012) reported that randomly selected cows displayed lower bias due to preferential treatment than elite cows. Wiggans et al (2011) presented an adjustment procedure on Mendelian sampling part of genotyped cows, which yielded a 3.5\% improvement in reliability over parent average for predicted Holstein bulls compared to use the unadjusted PTA records of cows in the Holstein training population for yield traits.

**Conclusions**

Use of cow data in the training population is an attractive way to implement genomic selection in countries with no or limited progeny-tested bull data. However, preferential treatment of elite cows may result in bias and lower prediction accuracy. A large cow training population can mitigate the effect of preferential treatment on accuracy and bias of genomic predictions.

**List of References**

Figure 1. Pearson correlation between TBVs and DGVs where none, 5, 10 and 20 percent of elite cows received preferential treatment (PT). a) $h^2=0.30$ and b) $h^2=0.05$.

Group refers to the numbers of cows used in the training population.