STØTTET AF
mælkeafgiftsfonden
INTERBULL BULLETIN NO. 47. Nantes, France, August 23 – 25, 2013

Detection of Genomic Pre-Selection with Mendelian Sampling Variance Test

A.-M. Tyrisevä[1], E.A. Mäntysaari[1], J. Jakobsen[2], G.P. Aamand[3], J. Dürr[2], W.F. Fikse[4] and M.H. Lidauer[1]

[1]*MTT Agrifood Research Finland, Biotechnology and Food Research, Biometrical Genetics, Jokioinen, Finland*
[2]*Interbull Centre, Department of Animal Breeding and Genetics, SLU, Uppsala, Sweden*
[3]*NAV Nordic Cattle Genetic Evaluation, Aarhus, Denmark*
[4]*Dept. Animal Breeding and Genetics, SLU, Uppsala, Sweden*

Abstract

Genomic pre-selection of bull calves is potentially one additional source of bias in the international evaluations, given it has not been adequately accounted for in the national evaluations. The aim of this simulation study was to investigate whether it is possible to detect the effect of genomic pre-selection in the MS means or variances obtained by the MS-trend validation test. In total, 50 replicates were generated under control and genomic pre-selection schemes utilizing structures of the real data and pedigree from a medium size cow population during a 20-year time interval. All bulls in the last 10 birth year classes were assumed to be genomically pre-selected. Therefore, their Mendelian sampling terms were inflated with a value corresponding to the selection of best 10% of the genomically tested bull calves. After start of genomic pre-selection, both the true and estimated means of Mendelian sampling terms and breeding values rose sharply in bulls, although the effects of estimated means were more moderate. A clear decrease was found in true and estimated genetic variances of bulls, but the effect was temporary and thus hard to exploit. Daughters of genomically pre-selected bulls had higher true and estimated breeding values compared to the control scheme, only slightly elevated MS means and no effect in the genetic variances.

Introduction

Trends in genetic variance affect ranking of bulls in the international evaluations and lead to sub-optimal selection decisions. One typical source of bias is an inadequate or a lacking heterogeneous variance adjustment of the national evaluation models. In the era of genomics, an additional source of bias is the genomic pre-selection of bull calves (Patry and Ducrocq 2011, Vitezica *et al.* 2011). The pre-selected bulls are no longer a random sample of the progeny of their parents, which results into an inflated mean of the Mendelian sampling (MS) terms for these bulls. In that case, normal assumptions of the BLUP model do not hold true and methods such as a single-step should be used instead (e.g., Vitezica *et al.*, 2011). Theoretically, also the genetic variance reduces due to pre-selection (Falconer and Mackay 1996). The aim of this simulation study was to assess if the MS-trend validation test (Tyrisevä *et al.*, 2012) is sensitive enough to detect the effect of genomic pre-selection in the MS means or variances and thus a suitable test for the current evaluation systems.

Material and Methods

Data used for simulations

The field data set used for the simulations consisted of 754 600 Danish Holstein cows from 2000 herds and a 20-year time interval. The pedigree information included 1.2 million animals (Tyrisevä *et al.*, 2011). For simulations, only the herd and the pedigree structure were retained from the original data and an artificial trait was simulated. We generated one record for each cow under a model including a fixed herd effect and random additive genetic and residual effects. The estimate of heritability used was 0.25.

Design of the study

Two schemes were created: a control and a genomic pre-selection (GPS) scheme, with 50 replicates in each (Figure 1). For both schemes, a genetic trend was first generated in the data. In the GPS scheme this was followed by the creation of genomic pre-selection for

the bulls belonging to the most recent birth year classes.

Generating genetic trend

In order to generate a genetic trend, original observations were replaced by yearly increasing values carrying a desired annual trend (step 1). Records of animals having progenies were set missing to ensure that the average MS terms of parents would not be regressed towards yearly means. These pseudo observations were used to estimate breeding values (BVs) that are in synchrony with parent and progeny averages and expected yearly means of BVs (step 2). The created pseudo BVs and the pedigree information were then used to calculate the expected MS terms $E[\boldsymbol{\varphi}]$ for each animal (step 3):

$$E[\boldsymbol{\varphi}] = \mathbf{L}^{-1}\mathbf{u}, \qquad (1)$$

where $\mathbf{L}^{-1}$ is from the decomposition of the numerator relationship matrix $\mathbf{A}=\mathbf{LDL}^{T}$ and determines the flow of genes from parents to offspring and $\mathbf{u}$ is the vector of pseudo BVs. Thus, the genetic trend embedded in $\mathbf{u}$ was transmitted to the MS terms of selected animals. These MS terms were used to simulate true BVs and phenotypic observations (step 3). The estimated BVs were obtained by fitting the same model as used for the data generation to the simulated data (step 4). For more detailed information of the simulation method, see Mäntysaari *et al.* (2013). All the analyses were carried out using the MiX99 software package (Lidauer *et al.*, 2011).

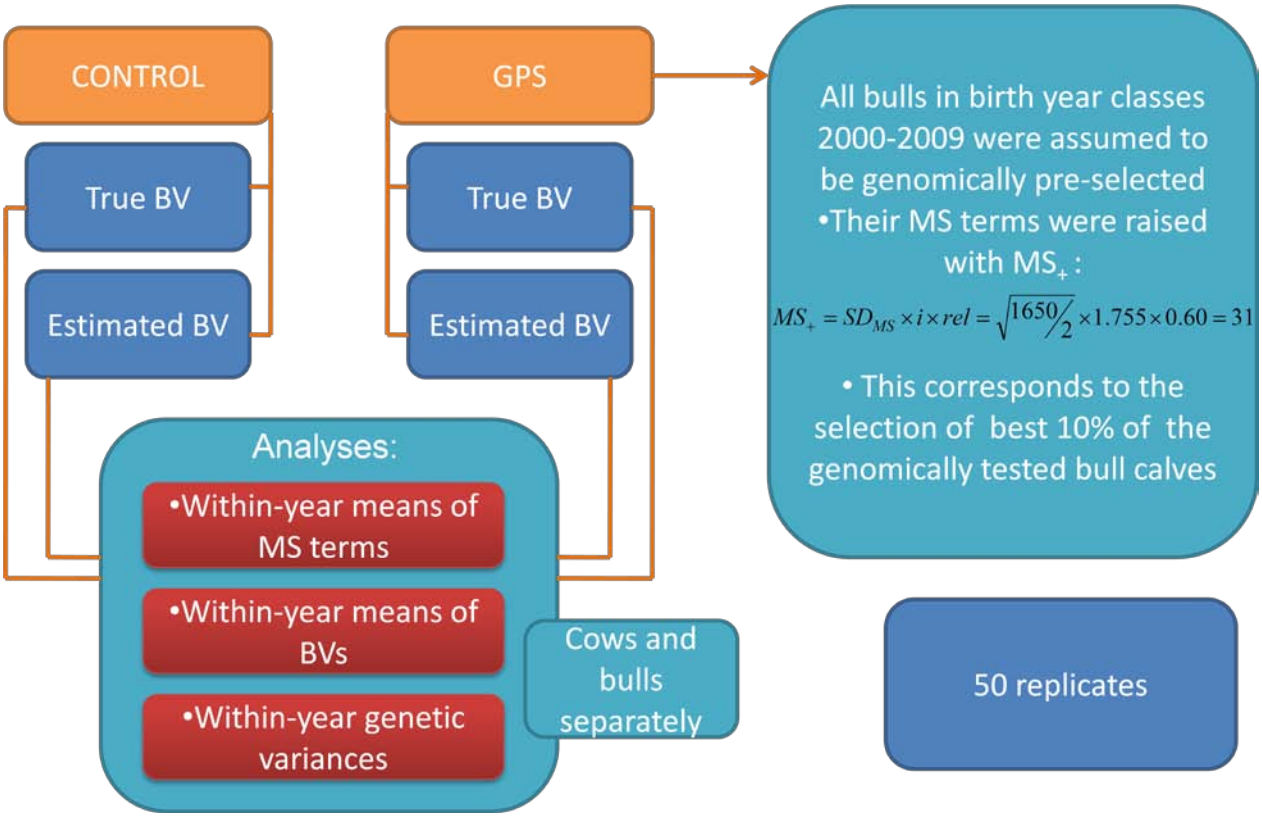

Figure 1. Design of the study.

Considering change in genetic variance

Under the non-zero expectation of MS terms, the MS terms no longer follow the normal distribution with the MS variance $\sigma_\phi^2 = d_{jj}\sigma_u^2$, in which d_{jj} is the diagonal of an animal j in the decomposition of $\mathbf{A}=\mathbf{LDL}^T$. In that case, the MS variance is increased with the variance of the expected MS terms, $\mathrm{var}(E[\phi])$, leading also into inflated variance of BVs. This was avoided by carrying out a variance reduction, so that $\sigma_\phi^2 = (1-k)d_{jj}\sigma_u^2$. According to the standard formula by Falconer and Mackay (1996), $k = i(i-x)$, where i is the selection intensity and x is a deviation of truncation point from the mean in standard deviation units. The selection intensity can be further formulated as $i = E[\phi]/\sigma_\phi$, in which σ_ϕ is a standard deviation of the MS terms. To find x for the given i is a non-linear approximation and computationally demanding. Since *(1-k)* is an exponential function of *i*, a satisfactory approximation can be obtained by a linear fit on its logarithmic value. The best fit was obtained using the following formula:

$$(1-k)_j = Exp(-1.18969|i| + 0.10805i^2), \quad (2)$$

where $|i|$ is the absolute value of i (Mäntysaari *et al.*, 2013).

Creating genomic pre-selection

All bulls born through 2000 to 2009 were assumed genomically pre-selected in the GPS scheme and their MS terms were inflated prior step 3 with a constant corresponding to the selection of best 10% of the genomically tested bull calves (Figure 1). The MS terms of all old bulls and cows were kept intact.

Analyses

True and estimated BVs were used to calculate within-year means of BVs, MS terms and genetic variances for cows and bulls separately. The latter two were obtained from the program designed to be used for the validation of MS-trend (Tyrisevä *et al.*, 2012).

Results and Discussion

Bulls

True and estimated yearly means of BVs, averaged over 50 replicates, were in practice identical under control scheme in bulls (Figure 2). After start of GPS, both the true and estimated BV means rose sharply, although the increase in the estimated means was more modest. This was because some of the total genetic progress was moved into the environmental trend instead.

Within-year MS means were very close to zero in bulls under the control scheme (Figure 3). The true MS means were identical to those under the GPS scheme until the GPS started. This led to the expected rise of yearly MS means to the level of +31, which was the constant added to the GPS bulls. Again, the effect of GPS on the estimated MS means was more modest, around 1/3 of that of the true means, but it clearly deviated from the zero expectation of the non-pre-selected population. At least two evident factors causing the difference between true and estimated means can be found. First, a part of the simulated genetic progress due to pre-selection went into the environmental effects. Second, the parental EBVs in the BLUP model tended to increase in consequence to the increase of the EBVs of their GPS sons. However, due to over-lapping generations, solutions found an equilibrium in which the GPS sons still had higher than zero MS terms.

Results for the within-year genetic variances are shown in Figure 4. When BVs were simulated, the genetic variance applied was 1650. The overall level of genetic variance was a bit lower in bulls. This was expected since the main force of selection is through bulls. Further, although the expected MS terms of bulls distributed around zero in simulations, most of the bulls had the MS term higher than zero.

At start of GPS, true genetic variance clearly decreased. However, after some years of GPS, it turned to increase back to its original level. This might be due to the Bulmer effect (Falconer and Mackay 1996). The estimated genetic variance declinedalso after

start of GPS, but in agreement with the results from the BV and MS means, the effect was not as strong as for the true genetic variance. Further, the estimated genetic variance returned quicker to the original level. One reason for that might be that the daughters of GPS bulls started to become bull dams.

Cows

Compared to the control scheme, means of true and estimated BVs started to increase in cows two years after the start of GPS in bulls, i.e., when the first GPS bulls became sires (Figure 5). However, the effect was small and the difference between the true and estimated means minor. The overall level of MS means was zero in cows (Figure 6). Only tiny deviation from the zero expectation could be seen for daughters of GPS bulls. Further, both the true and estimated genetic variances in both schemes were in practice identical with the variance applied when BVs were simulated (Figure 7). This also implies that the formula used for the variance correction (2) fitted well.

Conclusions

The effect of genomic pre-selection can be detected through within-year MS means that clearly deviated from the zero expectation in bulls – given that the GPS has not been adequately modeled in the national evaluations. The decline in the genetic variance was temporary and thus hard to exploit. The method applied to generate a genetic trend and a genomic pre-selection demonstrated to be useful for this kind of studies.

Acknowledgements

The study has been jointly funded by MTT, NAV and Interbull. The data for the analyses has been received from NAV.

References

Falconer, D.S. & Mackay, T.F.C. 1996. *Introduction to quantitative genetics*, fourth ed. Longman, Essex.

Lidauer, M.H., Matilainen, K., Mäntysaari, E.A. & Strandén, I. 2011. *Technical reference guide for MiX99,* release IV/2011. MTT Agrifood Research Finland, Jokioinen. http://www.mtt.fi/BGE/Software/MiX99.

Mäntysaari, E.A., Tyrisevä, A.-M., Lidauer, M.H. & Strandén, I. 2013. Simple method to generate animal breeding data sets with genetic trend. *Manuscript, submitted for publication in J. Anim. Breed. Genet.*

Patry, C. & Ducrocq, V. 2011. Evidence of biases in genetic evaluations due to genomic preselection in dairy cattle. *J. Dairy Sci. 94*, 1011-1020.

Tyrisevä, A.-M., Mäntysaari, E.A., Fikse, F. & Lidauer, M.H. 2011. Simulation study on Mendelian sampling variance tests. *Interbull Bulletin 44*, 57-61.

Tyrisevä, A.-M., Mäntysaari, E.A., Jakobsen, J., Aamand, G.P., Dürr, J., Fikse, W.F. & Lidauer, M.H. 2012. Validation of consistency of Mendelian sampling variance in national evaluation models. *Interbull Bulletin 46*, 97-102.

Vitezica, Z.G., Aguilar, I., Misztal, I. & Legarra, A. 2011. Bias in genomic predictions for populations under selection. *Genet. Res. 93*, 357-366.

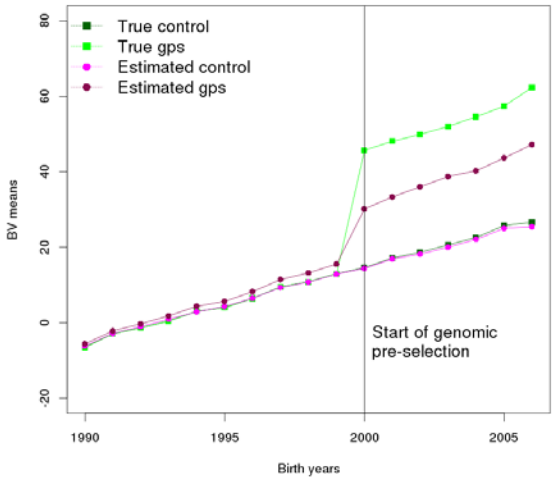

Figure 2. True and estimated means of BVs in bulls from control and genomic pre-selection (GPS) schemes. Within-year means were averaged over 50 replicates.

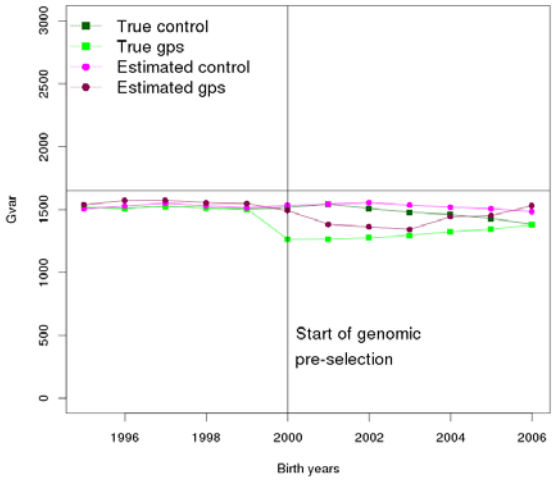

Figure 4. True and estimated within-year genetic variances in bulls from control and genomic pre-selection (GPS) schemes obtained from IB4. Within-year means were averaged over 50 replicates.

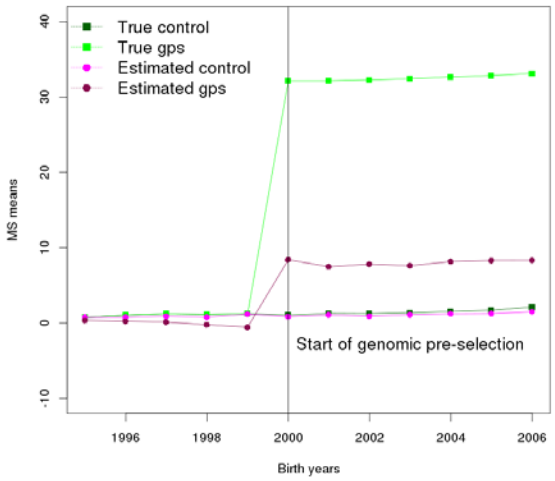

Figure 3. True and estimated MS means in bulls from control and genomic pre-selection (GPS) schemes. Within-year means were averaged over 50 replicates.

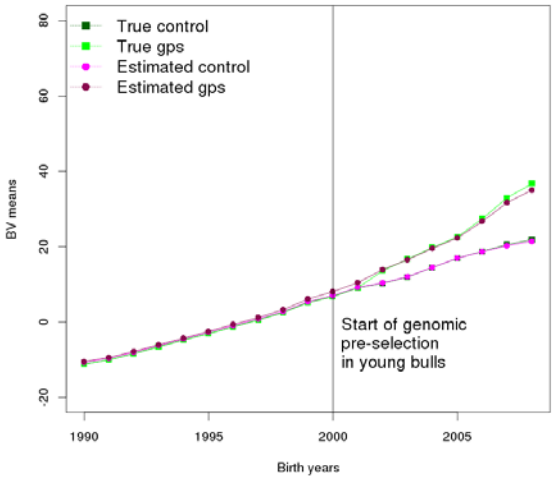

Figure 5. True and estimated means of BVs in cows from control and genomic pre-selection (GPS) schemes. Within-year means were averaged over 50 replicates.

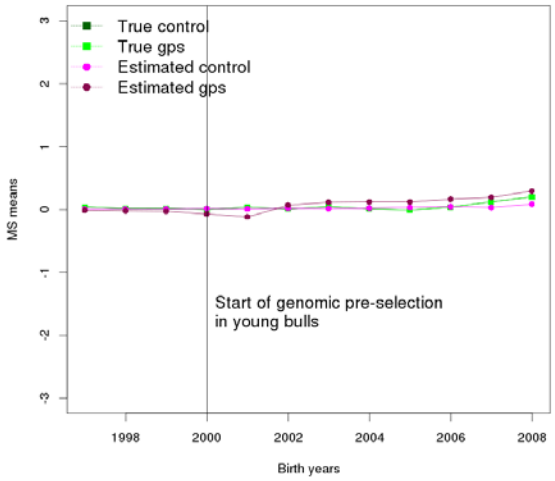

Figure 6. True and estimated MS means in cows from control and genomic pre-selection (GPS) schemes. Within-year means were averaged over 50 replicates.

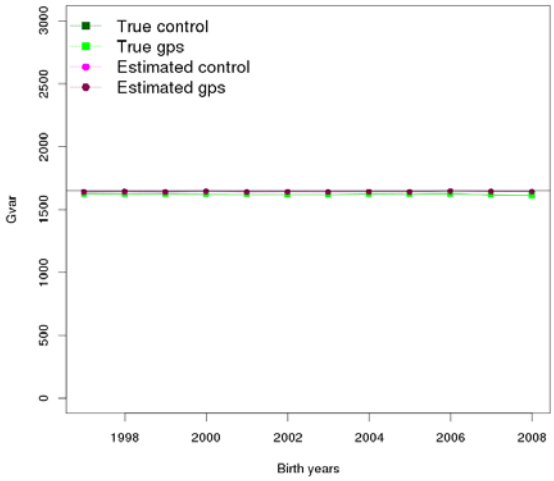

Figure 7. True and estimated within-year genetic variances in cows from control and genomic pre-selection (GPS) schemes obtained from IB4. Within-year means were averaged over 50 replicates.