Measuring Genomic Pre-Selection in Theory and in Practice

P.M. VanRaden[1] and J.R.Wright[1]

[1]*Animal Improvement Programs Laboratory, Agricultural Research Service, USDA, Beltsville, MD, USA*

Abstract

Potential biases from genomic pre-selection were estimated from actual selection and mating patterns of US Holsteins. Traditional models using only phenotypes and pedigrees do not adjust for average genomic merit of an animal's parents, progeny, mates, or contemporaries. Positive assortative mating of elite young bulls to elite genotyped females and dams of highly selected sons will become primary sources of bias in the next few years unless methods of adjustment are introduced. However, deregression can remove some biases such as by crediting the dam for only her own records and not for those of her selected progeny. Diagonals of genomic and pedigree relationship matrices, their inverses, and their differences were examined. The animal's own genotype provides a fourth source of information along with traditional information from parents, progeny, and own phenotype.

Key words: genomic evaluation, information sources, pre-selection bias, single-step model

Introduction

Genomic selection may bias an animal's traditional genetic evaluation if the average merit of its progeny, mates, or contemporaries differs from their merit estimated by the traditional animal model. Average merit may differ because only progeny with highest Mendelian sampling are selected and phenotyped, or because contemporaries do not get credit for their superior Mendelian sampling, or mating is nonrandom within herd based on genomic merit. Several studies simulated genomic pre-selection with random mating and estimated sizes of biases from progeny merit or contemporary merit (Liu et al., 2010; Nielsen et al., 2012; Patry and Ducrocq, 2011; Patry et al., 2013). Bias was larger with discrete than with overlapping generations. Vitezica et al. (2011) tested positive assortative mating, but selection and mating were based on estimated breeding value (EBV) rather than genomic EBV (GBV). Because of high demand and limited availability, the top young bulls might no longer be randomly sampled or mated to obtain their initial daughters.

Goals of this study are to 1) derive formulas for comparing sources of information contributing to genomic and traditional evaluations, 2) determine if genomic selection of mates or progeny is biasing traditional evaluations for elite new bulls and dams of bulls based on actual mating patterns and evaluations, and 3) discuss ability of deregression to remove some of these biases.

Methods

Genomic pre-selection should not bias single-step evaluations because all sources of information are used simultaneously (Aguilar et al., 2010), but most national and international evaluations continue to use multi-step procedures. In traditional evaluations, an animal's EBV is a weighted average of 3 sources of information: parent average (PA), own yield deviation (YD), and progeny contribution (PC), which is the average of twice its progeny EBV minus mates' EBV (VanRaden and Wiggans, 1991):

$$EBV = w_1\ PA + w_2\ YD + w_3\ PC,$$

where w_1, w_2, and w_3 sum to 1 and depend on numbers of known parents, records, progeny.

The GBV includes the animal's own genotype as a fourth source of information (GI) because mixed model equations (MME) for single-step models add variance ratio k times $\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}$ for genotyped animals along with $\mathbf{A}^{-1}$ for all animals. Solutions to MME are the same as traditional except that the off-diagonals of k ($\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}$) are multiplied by GBV of other genotyped animals and subtracted from the right-hand side for each genotyped animal. Then GI can be defined as

this sum of off-diagonals times GBV and divided by the diagonal of $\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}$. With this definition, GI measures information added by the animal's genotype beyond information already provided by the pedigree. The formula assumes that no new information is provided if the diagonal exactly equals 0.

Genotypes of parents, progeny, mates, more distant relatives, or animals competing in the same management groups also affect the animal's GBV. Let PA_g be the average of parents' GBV, YD_g be animal's yield deviation adjusted for GBV rather than EBV of management groupmates, and let PC_g be the average of twice progeny GBV minus mate's GBV (instead of EBV). Then:

$$GBV = w_1 PA_g + w_2 YD_g + w_3 PC_g + w_4 GI,$$

where weights w_1, w_2, and w_3 differ from those above because they sum to 1 with w_4 included. Similarly daughter yield deviation (DYD) can be computed and compared using EBV of mates and groupmates or using GBV (DYD_g; VanRaden, 2012), and reliability can be computed by summing daughter equivalents from 4 sources instead of 3.

Merit of mates was examined for two groups of Holstein bulls; those with daughter records and those whose daughters do not yet have records. Group 1 was the top 50 young genotyped Holstein bulls ranked on April 2010 net merit (NM$) that later had at least 50 daughters in their August 2013 evaluations. Group 2 was the top 50 genotyped bulls born in 2009 and 2010 that had already produced U.S. calves. Group 2 allows future bias to be forecast because the genomic pre-selection has already occurred and can be accurately measured when calves are born even if they do not have phenotypes yet.

Merit of progeny was examined for US Holstein dams of recently sampled US sons. Genomic selection did not occur before 2008 and cannot be measured in the last year until bulls get sampled, so dams were required have at least 1 genotyped and sampled son born January 2008 to April 2012. Counts of the total number of daughters and number of genotyped sons born before April 2012 were obtained for each dam, along with the selection differential of the sampled sons, computed by summing (GPTA – traditional PA) / number of sampled sons. Future bias in the dam's traditional evaluation was the sons' selection differential times DE from sampled sons divided by total DE from all of dam's traditional sources (not including genomic DE). Formulas from Table 1 of VanRaden and Wiggans (1991) were used to estimate a dam's DE from each source and assumed that with 25% heritability, 8.3 DE would come from dam's PA, 7.8 from her YD, 1.5 from each daughter, and 5.4 from each son.

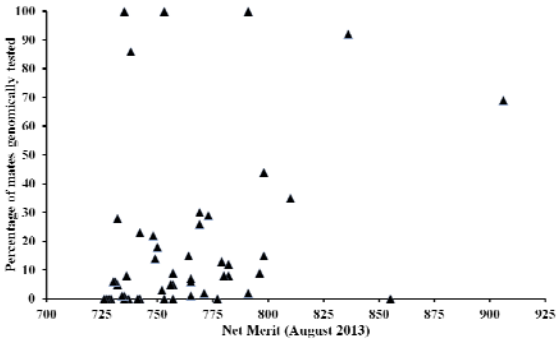
Deregression can help to remove biases present in the traditional EBVs if done jointly rather than one animal at a time. For example, if only the best sons of a dam are progeny tested, the dam's traditional EBV will be biased, but her deregressed evaluation will include direct credit only for her own records and those of any nongenotyped progeny. Each son gets credit for his progeny test minus any progeny genotyped, because those progeny will get credit for their own records directly. However, the deregressed y vector will still contain biases caused by incorrect credit for genomic merit of contemporaries or mates if the environmental effects in the model are also biased by pre-selection, because those are not re-estimated during deregression. If a bull is mated primarily to genomically selected mates, his traditional EBV will be biased, and this bias will be removed by deregression only if the progeny are genotyped. Fortunately, in the future perhaps most progeny will be genotyped, but some biases will remain because adjustments for genetic merit of contemporaries use the biased EBV in the traditional animal model.

Results and Discussion

Recently proven bulls (group 1) had little opportunity for bias in their traditional PTAs from genomic selection of mates because 45 of the 50 bulls had <20% of their mates genotyped. The other 5 bulls had from 27-46%, including the number 1 ranked young bull whose percentage of genotyped mates was highest at 46%. Of his 739 genotyped mates, 459 (62%) had been genotyped before the mating with this bull occurred, giving at least an opportunity for genomic pre-selection of mates. For group 2 bulls, the majority also had small percentages of genotyped mates, but 4

bulls had >90% genotyped mates and 2 others had >50% (Figure 1).

Figure 1. Genotyped mates (%) of the top young bulls (group 2) for net merit.



Differences between mates' genomic PTA (GPTA) minus traditional PTA for yield traits were small for nearly all bulls in both groups, indicating mostly random mating. The largest differences were for the bulls with the most genotyped mates, but were in the opposite direction as expected, with mates' GPTA a little lower than PTA on average. This may have occurred if breeders were not as confident of young bulls and instead mated their best GPTA females to progeny tested bulls instead of to young bulls in the early years of genomic selection. Differences for other traits were also small for most bulls, and combined to cause an expected bias > $30 net merit for only 2 bulls (Table 1).

Table 1. Future bias expected from genomic selection of mates of young bulls (group 2).

Trait	Mean	SD	Min	Max
Net Merit ($)	8	9	0	33
Protein (lb)	0	0	0	1
Prod. Life (mo)	0.1	0.1	0.0	0.5
Dtr. Preg. Rate (%)	0.1	0.1	0.0	0.3
SCS (log 2)	-0.01	0.01	-0.03	0.00
Final Score	0.02	0.03	-0.01	0.10
Udder Depth	0.03	0.04	-0.01	0.13

Other sources of bias could potentially affect EBV and GBV in the future. Elite GBV heifers may receive preferential treatment because of the very high prices paid. For traits with low to moderate heritability, a cow's own phenotype receives much less weight than the genomic information, reducing the impact of preferential treatment unless it affects many maternal relatives. For elite young bulls, limited distribution just after puberty gives owners an incentive to treat the first few daughters differently than the larger number of later daughters. However, little bias should occur this year or in the next couple of years because most daughters of the top ranking young bulls are from nearly random mates when measuring GBV – EBV differences.

The top 50 young bulls from 2010 average +560 NM$ in 2013, slightly better than the +542 for the top 50 bulls that already had daughter evaluations in 2010. The young bulls did have large decreases from the +713 predicted in 2010, whereas the daughter proven bulls decreased only a little from +563 in 2010. New adjustments have been applied since 2010 to reduce the overestimation. As selection differentials increase and more animals are genotyped, the advantages of young bulls should increase in the future. The correlation of GPTA NM$ from 2010 with 2013 was 0.82 compared to 0.66 for 2010 PA with 2013 GPTA, indicating good predictive accuracy.

Dams of highly pre-selected sons will have biased traditional PTA in the near future if adjustments are not made to account for the selection. An example cow had 2 selected sons from 12 genotyped, along with 7 daughters. The selection differential (GPTA – PA) for daughter pregnancy rate (DPR) was 1.6 for the 2 selected sons. Assuming 16.5 DE after each obtains progeny, plus 2.2 DE from each daughter, plus 58.2 DE from the dam's PA, plus 9.2 DE from the cow's own records, the future expected bias = 2 * 1.6 * 2 * 16.5 / [58.2 + 9.2 + 2 * 16.5 + 7 * 2.2] = 0.9, slightly more than half of the SD of true TA (1.70). Future biases from son selection were computed for several traits (Table 2).

Table 2. Future bias expected in traditional PTA for dams of genomically selected sons.

Trait	Mean	SD	Min	Max
Net Merit ($)	29	33	-124	156
Milk (lb)	27	132	-471	576
Fat (lb)	2	5	-16	21
Protein (lb)	1	3	-10	14
Prod. Life (mo)	0.3	0.5	-1.7	2.0
Dtr. Preg. Rate (%)	0.1	0.2	-0.9	0.9
SCS (log 2)	-.01	.04	-.22	.14

The new terms GI and w_4 are more difficult to interpret than the genomic adjustments to PA, YD, and PC. Diagonals of **G** and **A** were only moderately correlated when examined for 8,300 Brown Swiss animals, but diagonals of the 2 inverses and the difference between the 2 inverses were all highly correlated (Table 3). Diagonals of inverses are large for animals with many genotyped progeny, and are nearly uncorrelated to the animal's pedigree or genomic inbreeding on the original diagonal. More effort and examples are needed to show breeders not familiar with matrix algebra how GI contributes to evaluations.

Table 3. Correlations of diagonal elements of **G**, **A**, $\mathbf{G}^{-1}$, $\mathbf{A}_{22}^{-1}$, and $\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}$.

	G	**A**	$\mathbf{G}^{-1}$	$\mathbf{A}_{22}^{-1}$	$\mathbf{G}^{-1}$-$\mathbf{A}_{22}^{-1}$
G	1	.70	.05	.03	.06
A	.70	1	.02	-.02	.04
$\mathbf{G}^{-1}$	.05	.02	1	.98	.99
$\mathbf{A}_{22}^{-1}$	.03	-.02	.98	1	.94
$\mathbf{G}^{-1}$-$\mathbf{A}_{22}^{-1}$	.06	.04	.99	.94	1

Conclusions

To avoid biases, evaluations will need to adjust for GBV instead of EBV of progeny, mates, contemporaries, and parents. Current biases from mate pre-selection are small for recently proven bulls, but moderate biases will occur soon because the top young bulls are being mated to elite females with higher GBV than EBV. Dams of several highly selected young sons will have larger biases when the sons' daughters arrive, especially for traits with low heritability because sons contribute relatively more than daughters for those traits. Some of those biases can be removed by deregression, but other biases will grow as more animals are genotyped and pre-selection affects estimates of other factors in the model.

Acknowledgements

Phenotypes, genotypes, and pedigrees for this research were provided by the Council on Dairy Cattle Breeding and its member organizations. The authors thank Tom Lawlor for helpful suggestions.

References

Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J. Dairy Sci. 93, 743–752.

Liu, Z., Seefried, F., Reinhardt, F., and Reents, R. 2010. A simple method for correcting the bias caused by genomic pre-selection in conventional genetic evaluation. Interbull Bulletin 40, 184-188.

Patry, C. and Ducrocq, V. 2011. Accounting for genomic pre-selection in national BLUP evaluations in dairy cattle. Genet. Selection Evol. 43, 30.

Patry, C., Jorjani, H., and Ducrocq, V. 2013. Effects of a national genomic preselection on the international genetic evaluations. J. Dairy Sci. 96, 3272-3284.

Nielsen, U.S., Pösö, J., Madsen, P., Mäntysaari, E.A., Pedersen, J., Su, G., and Aamand, G.P. 2012. Effect of genomic preselection on the stability of EBV's from traditional BLUP procedure for production traits – a practical illustration. Presented at Interbull annual meeting, Cork, Ireland.

VanRaden, P.M. 2012. Avoiding bias from genomic pre-selection in converting daughter information across countries. Interbull Bull. 45, 29-33.

VanRaden, P.M., and Wiggans, G.R. 1991. Derivation, calculation, and use of national animal model information. J. Dairy Sci. 74, 2737–2746.

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