Modifying MACE to accommodate genomic preselection effects

P.G. Sullivan[a], E.A. Mäntysaari[b], G. DeJong[c], H. Benhajali[d]

[a]Lactanet, 660 Speedvale Ave, Guelph, Ontario, Canada, N1K 1E5
[b]Natural Resources Institute (Luke), PO Box 2, FI-00791 Helsinki, Finland
[c]CRV, PO Box 454, 6800 AL Arnhem, The Netherlands
[d]Interbull Centre, SLU, Box 7023, S-75007 Uppsala, Sweden

Abstract

Sire evaluations from MACE are used as input for national genomic evaluations. The MACE results are based on traditional evaluation models ignoring genotypes, at both the national and international levels. The exclusion of genotypes is to avoid a cyclical and repeated double-counting of genomic information between national and international systems. Ignoring the genotypes, however, has the consequence of introducing bias in the MACE results, because the effects of genomic preselection are not included in the MACE estimated breeding values of genomically preselected sires. The bias problem is especially relevant for most recent AI bulls, the young sires of most interest in current breeding programs. Current and future methods are discussed, which could be used to reduce genomic preselection biases in MACE, while still generating suitable MACE proofs that can be used as input to national genomic evaluation systems without double-counting the genomic information.

Key words: international evaluation, MACE, genomics, selection bias, double-counting

Introduction

Prior to the era of genomics, within-family selection of young bulls was only possible after progeny-testing. The preselection of bulls to be progeny-tested was based on parent averages (PA) from the EBV of sires and dams for each bull. Thus, bulls were drawn from only the best families, while the preselection of bulls within each family was essentially random. The distributions of true Mendelian Sampling (MS) values, for new crops of bulls used in AI each year, always had an expected value of zero. Biases were likely to be small due to family preselection, because recorded pedigrees were relatively complete for AI bulls (Kennedy et al, 1988) and in situations where pedigrees were incomplete, the corresponding effects of selection were accounted for by adding unknown parent groups to the model (Westell et al, 1988).

Since 2008, within-family selection of young bulls has become feasible without a progeny test, using genotype-based genomic evaluations (e.g. VanRaden, 2008). The expected value of MS for newly selected AI bulls has thus moved away from zero, increasing in magnitude as both the accuracies of genomic evaluation and the intensities of genomic selection have increased. Variances of MS are also affected by genomic preselection, in addition to the selection effects on MS averages (Sullivan, 2018). Preselection effects on the distributions of MS for AI bulls are not accounted for in traditional models that ignore genotypes, even with complete pedigrees, or with unknown parent groups accounting for incomplete pedigrees.

Genomic selection has clear and strong impacts on the MS distributions of recent young bulls selected for AI. Most evaluation systems for national EBV and MACE do not account for genomic preselection effects, although there are techniques available to at least partially account for these (e.g. Patry et al, 2013). The purposes of the present paper were to review methods currently available, and to discuss newer methods that could be developed, to account for genomic preselection effects in traditional evaluation models. The ultimate goal is to reduce genomic preselection bias in both national EBV and MACE results.

Continuing relevance of MACE

MACE proofs are commonly used as input for national genomic evaluations, in order to increase reference population sizes, the

reliabilities of estimated SNP effects and animal GEBV, and to improve direct genomic comparisons among bulls and their young progeny from different countries. All countries can increase genomic reliabilities and international genomic comparisons by including MACE proofs of foreign bulls in genomic evaluations. For countries with relatively few local bulls, the MACE proofs of foreign bulls are an especially important source of input data for national genomic evaluations.

Genetic correlations among countries are routinely estimated by Interbull, as an integrated part of the routine MACE evaluation service. The country correlations estimated for MACE are used in MACE, GMACE, and SNP-MACE, and are publicly available for potential use in other systems as well, like regional genomic evaluations that predict correlated GEBV (or SNP effects) by country.

Genomic preselection bias in EBV

National evaluations are expected to be biased, if the genotypes used to preselect AI bulls are not included in the evaluation data (Henderson, 1984; Sorensen and Kennedy, 1984; Schaeffer et al, 1998; Patry and Ducrocq, 2011; Masuda et al, 2018). Excluding the genotypes is required, however, when computing national EBV for input to MACE. The national EBV must be "genomics-free" so that the MACE results are also "genomics-free", and thus suitable as input for national genomic evaluation systems.

Genomic preselection biases in MACE results are most likely to be observed in the predicted average differences between selection groups (e.g. recent genomic versus historical progeny-tested bulls), and in the variance of predicted sire breeding values within a selection group (e.g. within the most recent group of genomically preselected bulls). A variance bias does not affect rankings, but could adversely affect variances of national genomic evaluations computed from MACE proofs, and the estimates of sire variances and possibly country correlations that are used in MACE. The variance biases due to genomic preselection are in the MS estimates, while estimates of PA rankings could be biased both within and across the preselection groups.

Genetic groups in national EBV

Modeling preselection effects with genetic groups can reduce, but might not eliminate, the preselection biases. The relative contributions of PA versus MS in the estimated breeding value of an individual, and the correct partitioning of preselection effects between PA versus MS solutions can be difficult, because PA and MS effects are often confounded in the data (Fikse, 2014). Prediction error is also a concern for genetic groups representing genomic preselection effects, because many preselection groups will have relatively few members. The genetic differences between groups might not be large enough to overcome high prediction errors with small groups (Kennedy, 1981).

Genetic groups in MACE

Even if preselection effects are properly accounted for and the national EBV are unbiased, it is still necessary to update modeling within MACE. Simulation studies by Patry et al (2013) showed that even with unbiased national EBV as input, the MACE results are still biased when bulls with national EBV included in MACE are preselected, based on their genomic evaluations as young calves.

Definitions of genetic groups used currently in MACE are based only on pedigree, and the groupings do not differ among country-traits included in MACE. The same genetic groupings are also used in the models for deregression and for MACE evaluation. A different grouping strategy is needed to account for genomic preselection effects, because preselection intensities were not the same in all countries. Each country uses a different group of bulls locally, and the expected distributions of MS will therefore be different for each country. Deviations from group average, for the same bull, could be very different between two countries with different levels of preselection intensity.

It is unclear if groups for genomic preselection should be defined the same way in the deregression model as in the model used for MACE evaluation. There needs to be a balance between adjusting for heterogeneity of input data, which is due to different national methodologies being used, versus treating

observed patterns in data as indications of genomic preselection effects. Research will be required, and strong validation tests needed, to make objective comparisons among models and among different strategies for genetic grouping in MACE.

Programming for known-animal groups

Unknown parent groups are commonly included in genetic evaluation models, at both the national and international levels, for simple and complex genetic models, and for any number of traits. The implementation is very simple, following the rules of Westell et al (1988), but these simple rules to not apply for groups of known animals with known parents.

A different implementation is required to fit genetic groups for genomic preselection effects. At a national level, each country would need to update EBV software with new code that allows grouping of known animals with known parents (e.g. Quaas, 1988), and the software used for MACE evaluation would also require new code for these effects.

Other evaluation options

Genetic groups, although useful, might not provide a perfect solution. The effectiveness of genetic groups accounting for genomic preselection effects will vary among applications, and evaluation systems used by different countries. Genetic grouping is also just one possible approach for addressing genomic preselection effects in MACE.

A variety of new and different methods are being considered to reduce preselection biases in national EBV. Hyper-parameters could be defined, which account for preselection effects differently than genetic groups. Distributions of MS estimates could be constrained in some way, based on prior knowledge about the preselection that has occurred. Expected distributions of the true MS after selection, which are underlying the observed data from daughters of preselected AI sires, could be used as prior information. Data augmentation could be used, by adding pseudo-records for genotyped bulls that were not chosen for AI (e.g. Ducrocq and Patry, 2010), although double-counting of genomic information is more likely to occur in national genomic systems using MACE data, if genomic pseudo-records are included in MACE.

It is still unclear which methods will work best, or if modified animal-based evaluation methods will continue to work acceptably well for international sire comparisons. The use of MACE and/or GMACE could someday be replaced by SNP-based evaluation systems (e.g. Goddard et al, 2018). At present, however, Interbull and its member countries continue to rely on animal-based evaluation systems for a variety of reasons. These systems should be updated to better fit current data, which are strongly affected by genomic preselection (Schaeffer, 2018).

Applications of single-step genomic evaluation (e.g. Misztal et al, 2010) are becoming more common, where genotypes are included and the ssEBV might therefore be relatively free of genomic preselection bias. There is growing interest to derive genomic-free evaluations directly from these systems, as input data for MACE, and different approaches with single-step systems are being tested (e.g. Lourenco et al, 2015). Rather than review the many different approaches in detail, the relevant point for current discussion is that different approaches will likely be used by different countries, and Interbull can expect to see increased heterogeneity among the national data available for use in MACE.

Modifying MACE

Interbull needs to expect and plan for a more heterogeneous mixture of data as input to MACE, especially regarding MS distributions from EBV of different countries. The impacts of genomic preselection on MS distributions are not trivial, and there is currently no standard methodology to account for genomic preselection effects in the input data provided for MACE. Expanding the use of genetic groups should be considered as a first step within the MACE system, but additional or different changes might also be needed. Data transformations could be used, for example, to reduce heterogeneity among the input data sets provided by countries for MACE, and the MACE results back-transformed to ensure international results from MACE continue to be directly comparable with national data. A simple example of data transformation would

be a genetic trend adjustment to correct for expected preselection bias in the most recent trends for national EBV.

Presently, the national evaluations provided to Interbull for MACE are limited to males, while genomic preselection usually considers both males and females, and especially the dams of young bulls. Interbull would have a larger scope of options for data transformation, and for studying the impacts of genomic preselection on MS distributions and MACE, if countries provided national evaluations of bull dams, in addition to the current data being provided only for males.

Summary

Genomic preselection of AI sires alters the distributions of both true and estimated MS deviations. The MS distributions have reduced variance, and means that deviate significantly from zero. In national EBV and MACE models, the partitioning of PA and MS contributions to animal breeding values is based on the assumption that MS values are always drawn randomly. This assumption is no longer valid after genomic preselection has occurred. The traditional models should be updated with new expectations for the MS distributions of most recent AI sires.

References

Ducrocq, V. & Patry, C. 2010. *Combining genomic and classical information in national BLUP evaluation to reduce bias due to genomic pre-selection.* Interbull bulletin 41:33-36.

Fikse, F. 2014. *Can a model with genetic groups for Mendelian sampling deviations correct for pre-selection bias?* Proc. 10th WCGALP, Vancouver, Canada.

Goddard, M.E., Jighly, A. & Benhajali, H. 2018. *SNPMace – A meta-analysis to estimate SNP effects by combining results from multiple countries.* Interbull bulletin. 54:1-6.

Henderson, C.R. 1984. *Applications of linear models in animal breeding. Chapter 33.* University of Guelph, Guelph, Canada.

Kennedy, B.W. 1981. *Bias and mean square error from ignoring genetic groups in mixed model sire evaluation.* J. Dairy Sci. 64:689-697.

Kennedy, B.W., Schaeffer, L.R. & Sorensen, D.A. 1988. *Genetic properties of animal models.* J. Dairy Sci. 71(Suppl. 2):17-26.

Lourenco, D.A.L, Fragomeni, B.O., Tsuruta, S., Aguilar, I., Zumbach, B., Hawken, R.J., Legarra, A. & Misztal, I. 2015. *Accuracy of estimated breeding values with genomic information on males, females, or both: an example on broiler chicken.* Gen. Sel. Evol. 47:56.

Masuda, Y., VanRaden, P.M., Misztal, I. & Lawlor, T.J. 2018. *Differing genetic trend estimates from traditional and genomic evaluations of genotyped animals as evidence of preselection bias in US Holsteins*. J. Dairy Sci. 101:5194-5206.

Misztal, I., Aguilar, I., Legarra, A., Tsuruta, S., Johnson, D. L. & Lawlor, T. J. 2010. *A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation.* In: Proceedings of 9th World Congress on Genetics Applied to Livestock Production, Leipzig, Germany, paper 0050.

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

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

Quaas, R.L. 1988. *Additive genetic model with groups and relationships.* J. Dairy Sci. 71:1338-1345.

Schaeffer, L.R. 2018. *Necessary changes to improve animal models.* J. Anim. Br. Genet. 135:124-131.

Schaeffer, L.R., Schenkel, F.S. & Fries, L.A. 1998. *Selection based on animal model evaluation.* Proc. 6th WCGALP, Armidale, Australia, 25:501-508.

Sorensen, D.A. & Kennedy, B.W. 1984. *Estimation of genetic variances from unselected and selected populations.* J. Anim. Sci. 59:1213.

Sullivan, P.G. 2018. *Mendelian Sampling variance tests with genomic preselection.* Interbull bulletin. 54:1-4.

VanRaden, P. 2008. *Efficient methods to compute genomic predictions.* J. Dairy Sci. 91:4414–4423.

Westell, R.A., Quaas, R.L & Van Vleck, L. 1988. *Genetic groups in an animal model.* J. Dairy Sci. 71:1310-1318.