J. Dairy Sci. 101:1–9
https://doi.org/10.3168/jds.2017-13527

Detection of evaluation bias caused by genomic preselection

A.-M. Tyrisevä,*[1] E. A. Mäntysaari,* J. Jakobsen,† G. P. Aamand,‡ J. Dürr,§ W. F. Fikse,#[2] and M. H. Lidauer*
*Natural Resources Institute Finland (Luke), Green Technology, Biometrical Genetics, 31600 Jokioinen, Finland
†Norwegian Association of Sheep and Goat Breeders, 1431 Ås, Norway
‡NAV Nordic Cattle Genetic Evaluation, 8200 Aarhus, Denmark
§Council on Dairy Cattle Breeding, Bowie, MD 20716
#Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, 75007 Uppsala, Sweden

ABSTRACT

The aim of this simulation study was to investigate whether it is possible to detect the effect of genomic preselection on Mendelian sampling (MS) means or variances obtained by the MS validation test. Genomic preselection of bull calves is 1 additional potential source of bias in international evaluations unless adequately accounted for in national evaluations. Selection creates no bias in traditional breeding value evaluation if the data of all animals are included. However, this is not the case with genomic preselection, as it excludes culled bulls. Genomic breeding values become biased if calculated using a multistep procedure instead of, for example, a single-step method. Currently, about 60% of the countries participating in international bull evaluations have already adopted genomic selection in their breeding schemes. The data sent for multiple across-country evaluation can, therefore, be very heterogeneous, and a proper validation method is needed to ensure a fair comparison of the bulls included in international genetic evaluations. To study the effect of genomic preselection, we generated a total of 50 replicates under control and genomic preselection schemes using the structures of the real data and pedigree from a medium-size cow population. A genetic trend of 15% of the genetic standard deviation was created for both schemes. In carrying out the analyses, we used 2 different heritabilities: 0.25 and 0.10. From the start of genomic preselection, all bulls were genomically preselected. Their MS deviations were inflated with a value corresponding to selection of the best 10% of genomically tested bull calves. For cows, the MS deviations were unaltered. The results revealed a clear underestimation of bulls' breeding values (BV) after genomic preselection started, as well as a notable deviation from zero both in true and estimated MS means. The software developed recently for the MS validation test already produces yearly MS means, and they can be used to devise an appropriate test. Mean squared true MS of genomically preselected bulls was clearly inflated. After correcting for the simulated preselection bias, the true genetic variance was smaller than the parametric value used to simulate BV, and also below the variance based on the estimated BV. Based on this study, the lower the trait's heritability, the stronger the bias in estimated BV and MS means and variances. Daughters of genomically preselected bulls had higher true and estimated BV compared with the control scheme and only slightly elevated MS means, but no effect on genetic variances was observed.

Key words: genomic preselection, Mendelian sampling, evaluation bias, data validation

INTRODUCTION

Unbiased comparison of dairy bulls among countries ensures efficient genetic progress of herds for dairy farmers and fair trade for breeding companies selling the best-ranking bulls on the global market. The bulls included in Interbull international evaluations receive EBV in each participating country's own unit, scale, and base, which enables comparisons of national and foreign sires. Interbull international evaluations employ multiple across-country evaluations (**MACE**), with national evaluation results, such as breeding values (**BV**) and transmitting abilities (**TA**), as input. The MACE has been found to be sensitive to the quality of national evaluation models and, consequently, to the BV or TA that countries provide for inclusion in Interbull international sire evaluations (e.g., Ducrocq et al., 2003). As observed in many studies, MACE is exposed to and affected by biased genetic trends and biased genetic variance trends in the national evaluations. Bulls from

---

Received July 19, 2017.
Accepted December 10, 2017.
[1] Corresponding author: anna-maria.tyriseva@luke.fi
[2] Current address: Växa Sverige, 75105 Uppsala, Sweden.

countries that overestimate their genetic trends, as well as bulls from birth year classes with inflated genetic variances, gain unfair advantage (Weigel et al., 1996; Van Doormaal et al., 1999; Gengler et al., 2000; Miglior et al., 2002; Ducrocq et al., 2003).

The era of genomic selection has given rise to a new source of bias: namely, bias due to genomic preselection of young bulls unless adequately accounted for in national genetic evaluations. The BV of each animal is estimated by mixed model equations (**MME**), including the parental average, the animal's own yield deviation, and the contribution of its offspring. Sorensen and Kennedy (1984) found that selection caused no bias in traditional genetic evaluation, provided that the data and pedigree of all animals were included; however, that is not the case with genomic preselection if data on culled bulls are excluded. Genomic enhanced breeding values (**GEBV**) also become biased if calculated using a multistep procedure (Patry and Ducrocq, 2011b; Vitezica et al., 2011). This procedure first involves traditional BV estimation to obtain pseudo-observations for the genotyped animals, utilizing their phenotyped relatives, after which these pseudo-observations are combined with marker information to determine the genotyped animals' GEBV (e.g., VanRaden, 2008; Hayes et al., 2009). The pseudo-observations are no longer a random sample of Mendelian sampling (**MS**) and the effect of selection cannot be accounted for, as only a subset of the available data are used to obtain GEBV (Ducrocq and Liu, 2009; Vitezica et al., 2011).

To solve the inherent source of bias introduced by genomic preselection, researchers have worked hard to reap the benefits of genomics while at the same time overcoming the challenges it brings. Indeed, new methods have been developed that are unbiased or at least reduce bias. For instance, one of the suggested approaches transforms GEBV into weighted deregressed performances, which are then used together with the original phenotypes in BLUP evaluations, allowing the inclusion of culled bulls (Ducrocq and Liu, 2009; Patry and Ducrocq, 2011a). The most elegant option, although computationally more demanding, is a single-step method (e.g., Aguilar et al., 2010; Christensen and Lund, 2010), which is an extension of traditional MME. It adds a new source of information, genotypes, to the traditional data from the animal itself, its parents, and its offspring. For this method, genotyped as well as nongenotyped animals are included in the model, and unbiasedness is achieved by including culled bulls. Development efforts to ease the computational burden of the single-step method are ongoing in several countries, and various solutions have already been suggested (e.g., Misztal et al., 2014; Taskinen et al., 2017).

Currently, 18 of the 31 countries (60%) participating in international bull evaluations have already adopted genomic selection in their breeding schemes: some of them only for production traits, some for a wider variety of traits. So far, only 2 of the 18 countries have implemented the single-step approach, whereas the majority rely on the original multistep approach (Interbull, 2017). This makes the data sent for MACE from different countries very heterogeneous and creates a real risk of bias due to genomic preselection.

In a simulation study with 3 countries, Patry et al. (2013) observed that genomically preselected bulls from countries that send biased data for international evaluations were penalized both for the biased trait and for correlated traits of other countries. In the context of MACE, the correlated trait, although the same biological trait, is treated as a different trait by various countries, allowing less-than-unity genetic correlation. The penalizing effect was to some extent also transmitted to the relatives of the selected bulls. Patry et al. (2013) further showed that failure to account for genomic preselection in national evaluations had a more severe effect than provision of incomplete (excluding culled animals) but unbiased data for international evaluations. The current heterogeneous situation among participating countries makes it very difficult to predict the overall effects of genomic preselection on the accuracy of MACE. VanRaden and Wright (2013) pointed out that bias is likely to increase in the near future, as elite young bulls are increasingly mated to elite genotyped cows or to cows with many good sons. Therefore, methods to account for genomic selection in national evaluations are needed, as well as for proper validation methods to detect possibly biased EBV.

A new test to validate the consistency of MS variance was recently developed and has been approved among the compulsory validation tests for countries participating in international sire evaluations (Tyrisevä et al., 2018). A tailored program for conducting the analyses is also available. The validation procedure estimates within-year genetic variances utilizing information on animals' MS values and also tests for a possible trend and outliers in the estimated variances.

The aim of our simulation study was to establish whether it is possible to use the new MS validation test to detect the bias caused by genomic preselection of young bulls either from the estimates of within-year MS means or variances. Theoretically, the variance of true breeding values of selected individuals is expected to decrease as a result of genomic preselection (Falconer and Mackay, 1996); further, the mean of MS deviations for selected individuals is expected to differ from zero (Patry and Ducrocq, 2011b).